OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: August 13, 2002, 17:36:08 ; Search time 1833.38 Seconds

Title: US-09-761-580-1_COPY_800_1800
Perfect score: 1001   (without alignments)
Sequence: 1 atatatacagcattgctgac.........cagttggagccagtggctt 1001
                                                            11425.598 Million cell updates/sec
Scoring table: IDENTITY_NUC
                      Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046326829 residues

                      1l number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*

             1: gb_ba:*
             2: gb_htg:*
             3: gb_in:*
             4: gb_om:*
             5: gb_ov:*
             6: gb_pat:*
             7: gb_ph:*
             8: gb_pl:*
             9: gb_pr:*
            10: gb_ro:*
            11: gb_sts:*
            12: gb_sy:*
            13: gb_un:*
            14: gb_vi:*
            15: gb_ba:*
            16: em_fun:*
            17: em_hum:*
            18: em_in:*
            19: em_mu:*
            20: em_om:*
            21: em_or:*
            22: em_ov:*
            23: em_pat:*
            24: em_ph:*
            25: em_pl:*
            26: em_ro:*
            27: em_sts:*
            28: em_un:*
            29: em_vi:*
            30: em_htg_hum:*
            31: em_htg_inv:*
            32: em_htg_other:*
            33: em_htgo_inv:*

SUMMARIES

Result   %
No.   Score   Query Match Length DB ID   Description
---- -----   ------------------   --   -----------

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

No.	Score	% Query Match	Length DB	ID			Description
1	1001	100.0	2583	6		AX151744	AX151744 Sequence
2	1001	100.0	2583	9		HSPDCE2	Y00978 Human mRNA
3	1001	100.0	2696	9		AK057299	AK057299 Homo sapi
4	961	96.0	2540	9		HUMIGMBC	J03866 Homo sapien
5	871.4	87.1	2543	4		AB036739	AB036739 Sus scrof
6	768.8	76.8	1700	10		AY044265	AY044265 Mus muscu
7	747.8	74.7	2266	10		RATPDCE2	D10655 Rattus norv
8	735.8	73.5	1370	6		A05167	A05167 Synthetic P
9	735.8	73.5	1370	6		A05167	A05167 Synthetic P
10	735.8	73.5	1370	10		RATPBCAA	D00092 Rattus norv
11	701	70.0	966	6		A05169	M16075 Rat primary
12	655.8	65.5	660	6		HSDHACTYL	A05169 H.sapiens m
13	264	26.4	2038	6		BC003202	X13822 H.sapiens m
14	223	22.3	319	6		AX305911	BC003202 Mus muscu
15	210	21.0	1833	6		DDU06634	AX305911 Sequence
16	205.8	20.6	40897	8		SPCC7764	U06634 Dictyostell
17	202.2	20.2	1620	8		AY033001	SPCC7764 S.pombe c
18	200.8	20.0	2001	8		AF367302	ALO23595 Arabidops
19	186.8	18.7	1981	8		AF135014	AY033001 Arabidops
20 c	178.4	17.8	27797	3		CEF23812	AF367302 Arabidops
21	170.2	17.0	40961	8		CBRG27D19	AF135014 Zea mays
22	169.2	16.9	172	6		AX341497	Z77659 Caenorhabdi
23	168.4	16.8	8238	1		ZMPDHBGEN	CBRG27D19 Caenorhab
24	167	16.7	161920	2		AP002502	AX341497 Sequence
25 c	156.8	15.7	175419	2		AP000788	X93605 Zymonas m
26	167	16.7	175419	2		AP000788	AP002502 Homo sapi
27	167	16.7	187516	2		AP000648	AP000788 Homo sapi
28	162	16.2	11199	1		AE006633	AP000788 Homo sapi
29 c	162	16.2	279110	1		RPXX03	AE006633 Ricketts1
30	157.6	15.7	1744	9		HSPRGTX	AJ235272 Ricketts1
31 c	157.6	15.7	1949	9		U79296	Y13145 Homo sapien
32	157.6	15.7	2320	9		AF001437	U79296 Human dihyd
33 c	156.8	15.7	175419	2		AP000788	AF001437 Homo sapi
34	156	15.6	2250	9		BC010389	AP000788 Homo sapi
35	156	15.6	2365	9		HSU82328	BC010389 Homo sapi
36	141.2	14.1	74587	2		AC025500	U82328 Homo sapien
37	141.2	14.1	110931	2		AC027079	AC025500 Mus muscu
38	140.8	14.1	62080	2		AC094189	AC027079 Mus muscu
39	134.8	13.5	9353	2		AE009644	AC094189 Rattus no
40 c	127.2	12.7	2175	8		ATDHPATF	AE009644 Brucella
41	127.2	12.7	2274	8		AF066080	Z46230 A.thaliana
42 c	125.8	12.6	41309	2		AC017138	AF066080 Arabidops
43	125.8	12.6	159350	3		AC008354	AC017138 Drosophil
44	125.8	12.6	257224	3		AE003617	AC008354 Drosophil
45	124.6	12.4	660	8		CNS01BMX	AE003617 Drosophil
							AL114433 Botrytis

ALIGNMENTS

RESULT 1
AX151744
LOCUS   AX151744   2583 bp   DNA   linear   PAT 22-JUN-2001
DEFINITION   Sequence 1 from Patent EP1118679.
ACCESSION   AX151744
VERSION   AX151744.1   GI:1453673
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2583)
  AUTHORS   Smith,J.C., Anand,R. and Morten,J.E.
  TITLE   Method for diagnosing polymorphisms in the human pdh e2 gene
  JOURNAL   Patent: EP 1118679-A 1 25-JUL-2001;
             AstraZeneca AB (SE)
FEATURES   Location/Qualifiers
     source   1..2583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   742 a   562 c   564 g   715 t
ORIGIN

Query Match   100.0%;   Score 1001;   DB 6;   Length 2583;
Best Local Similarity 100.0%;   Pred. No. 2e-234;
Matches 1001;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy     1 atatatcagcatttgctgactataggccaaccgaagtaacagatttaaaaccacaagtgc   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   800 ATATATCAGCATTTGCTGACTATAGGCCAACCGAAGTAACAGATTTAAAACCACAAGTGC  859

Qy    61 caccacctaccccaccccccgggttggccgctggtcctccacctccagccttagctccta  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   860 CACCACCTACCCCACCCCCGGGTTGGCCGCTGGTCCTCCACCTCCAGCCTTAGCTCCTA   919

Qy   121 caccttcagcaccctgcccagctactcctgtcggaccaaaggaaggggtgttgttagcc   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   920 CACCTTCAGCACCCTGCCCAGCTACTCCTGTCGGACCAAAGGGAAGGGGTGTTGTTAGCC  979

Qy   181 ctcttgcaagaagtggcagtagagaaggggattgatcttacacaagtaaaaggaacag   240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   980 CTCTTGCAAGAAGTGGCAGTAGAGAAAGGGGATTGATCTTACACAAGTAAAAGGGACAG  1039

Qy   241 gaccaagatggtagaatcaccaagagaagaggatatcgactccttttgtgtcctagtaaagttgctc  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1040 GACCAGATGGTAGAATCACCAAGAGAAGGATATCGACTCCTTTTGTGTCCTAGTAAAGTTGCTC  1099

Qy   301 ctgctccggcagctgtgtgtcctccccacaggtcctggaatggcaccagttcctacaggtg  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1100 CTGCTCCGGCAGCTGTGTGTCCTCCCCACAGGTCCTGGAATGGCACCAGTTCCTACAGGTG  1159

Qy   361 tcttcacagatatcccaaatcaacaacattcgcggttattcgcacacagcgattaatgcaat  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1160 TCTTCACAGATATCCCAAATCAGCAACATTCGCGGTTATTCGCACACAGCGATTAATGCAAT  1219

Qy   421 caaagcaaccataccctccattattaccttctcatcgatgtaaatatgggagaagttttgt  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1220 CAAAGCAACCATACCCTCCATTATTACCTTTCTCATCGATGTAAATATGGGAGAAGTTTTGT  1279

Qy   481 tggtacggaagaactcaataagagatattagaaggagagaacaaaatttctgtcaatgact  540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1280 TGGTACGGAAGAACTTAATAAGAGATATTAGAAGGAGAGAACAAAATTTCTGTCAATGACT  1339

Qy   541 tcatcataaaagctcagcttggcatgttaaaagttcccgaagcaaatctttctctttgga  600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1340 TCATCATAAAAGCTTCAGCTTGGCATGTTAAAAGTTCCCGAAGCAAATTCTTTCTTGGA   1399

Qy   601 tggacacagttatcaagacaaaatcatgtttgtcgtcagtgttcagtactcctg   660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1400 TGGACACAGTTATAAGACAAAAATCATGTTTGTTGTCAGTGTTCAGTACTCCTG   1459

Qy   661 caggactcatcacacctattgtgttaatgcacatataaaaggagtggaaacccattgcta  720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1460 CAGGACTCATCACACCTATTGTGTTAATGCACATATAAAAGGAGTGGAAACCCATTGCTA  1519

Qy   721 atgatgttgttcttttagcacaacaatttggaattagaagaatttctctgcta   780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1520 ATGATGTTGTTCTTTTAGCACAACAATTTGGAATTAGAAGAATTTCTCTGCTA   1579

Qy   781 agggtggcactcttacgatctccaaatttaggaattgttggaattacagagaaatcc   840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1580 AGGGTGGCACTTTTACGATCTCCAAATTTAGGAATTGTTGGAATTACAGGGAAATCC   1639

Qy   841 ttattaaccacctcagcacatgtatttggccatgttgctcagagagataaactggtcc   900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1640 TTATTAACCCACCTCAAGCACATGTTGTATTTGGCCATGTTGCTCAGAGAGGATAAACTGGTCC  1699

Qy   901 ctgcagataatgaaaaagggttttgatgtggctagcatgatgagtgtctgttacttcagtgtg  960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1700 CTGCAGATAATGAAAAAGGGTTTGATGTGGCTAGCATGATGAGTGTCTGTTACACTCAGTGTG  1759

Qy   961 atcaccggggtggtgatggagagaagttggagccccagtggctt  1001
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1760 ATCACCGGGTGGTGGATGGAGCAGTTGGAGCGCCAGTGGCTT  1800
```

RESULT 2

HSPDCE2

LOCUS   HSPDCE2   2583 bp   mRNA   linear   PRI 28-JUL-1995
DEFINITION Human mRNA for dihydrolipoamide acetyltransferase (PDC-E2) (EC 2.3.1.12).
ACCESSION   Y00978 X13969
VERSION   Y00978.1 GI:35359
KEYWORDS   dihydrolipoamide acetyltransferase.
SOURCE   human.
ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2583)
  AUTHORS   Patel,M.S.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-FEB-1989) Patel,M.S., Case Western Reserve
    University, Dept. of Biochemistry, School of Medicine, 2119
    Abington Road, Cleveland, Ohio USA 44106
REFERENCE   2 (bases 1 to 2583)
  AUTHORS   Thekkumkara,T.J.; Ho,L.; Wexler,I.D.; Pons,G., Liu,T.C. and
    Patel,M.S.
  TITLE   Nucleotide sequence of a cDNA for the dihydrolipoamide
    acetyltransferase component of human pyruvate dehydrogenase complex
  JOURNAL   FEBS Lett. 240 (1-2), 45-48 (1988)
  MEDLINE   89052894
COMMENT   The sequence overlaps with that reported by Coppel et. al. in Proc.
    Natl. Acad. Sci. USA 85:7317-7321(1988) J03866. Data kindly
    reviewed (29th March 1989) by Patel M.S.
FEATURES   Location/Qualifiers
  source   1..2583
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /tissue_type="liver"
                  /clone_lib="lambda gt11"
  CDS   <1..1848
                  /codon_start=1
                  /product="PDC-E2 precursor (AA -54 to 561)"
                  /protein_id="CAA68787.1"
                  /db_xref="GI:35360"
                  /db_xref="SWISS-PROT:P10515"
                  /translation="RVTSRSGPAPARNSVTGYGGVRALCGWTPSSGATPRNRLLLQ
           LLGSPGRRYYSLPPHQKVPLPSLSPTMQAGTIARWEKKKEGDKINEGDLIAEVETDKAT
           VGFESLEECYMAKILVAEGTRDVPIGAILCIITVGKPEDIEAPKNYTLDSSAAPTPQAA
           PAPTPAATPASSPPTSAQAPGSSYPPHMQVLLPALSPTMTMGTVQRMEKKVGEKLSEGD
           LLAEIETDKATIGFEVQEEGYLAKIIVPEGTRDVPLGTPLCIIVEKEADISAFADYRP
           TEVTDLKPQVPPPTPPPVAAVPPTPQPLAPTPSAPCPATPAGPKGRVFVSPLAKKLAV
           EKGIDLTQVKGTGPDGRITKKDIISFVPSKVAPAAAVPPGPGMAPVPTGVFTDIP
           ISNIRRVIAQRLMQSKQTIPHYLYISIDVNMGEVLLVRKELNKILEGRSKISVNDFIIK
           ASALACLKVPEANSSWMDTVIRQNHVDVSVAVSTPAGLLTPIVFNAHIKGVETIAND
           VVSLATKAREGKLQPHEFQGGTFTISNLGMFGIKNFSAIINPQACLIAIGASEDKLV
           PADNEKGFDVASMMSVTLSCDHRVVDGAVGAQWLAEFRKYLEKPITMLL"
  transit_peptide <1..162
  mat_peptide   163..1845
                  /product="mature PDC-E2 (AA 1-561)"
BASE COUNT   742 a   562 c   564 g   715 t
ORIGIN

Query Match   100.0%;   Score 1001;   DB 9;   Length 2583;
Best Local Similarity 100.0%;   Pred. No. 2e-234;
Matches 1001;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy     1 atatatcagcatttgctgactataggccaaccgaagtaacagatttaaaaccacaagtgc   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   800 ATATATCAGCATTTGCTGACTATAGGCCAACCGAAGTAACAGATTTAAAACCACAAGTGC  859

Qy    61 caccacctacccccccccccgggttggccgctggtcctccacctccagccttagctccta  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   860 CACCACCTACCCCACCCCCGGTGGCCGCTGGTCTTCCACCTCCAGCCTTAGCTCCTA    919

Qy   121 caccttcagcaccctgcccagctactcctgtcggaccaaaggaaggggtgttgttagcc  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   920 CACCTTCAGCACCCTGCCCAGCTACTCCTGTCGGACCAAAGGGAAGGGGTGTTGTTAGCC  979
```

```
Db  920  CACCTTCAGCACCCTTGGACCTACTCCCTGCTGCTGGACCAAGCTCCAAAAGGGAAGGGTGTTGTTAGCC 979
Qy  181  cctcttgcaaagaagttggcagtagagaaaggggatctcttaccacaagtaaaaggggacag 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  980  CTCTTGCAAAGAAGTTGGCAGTAGAGAAAGGGGATTGATCTTACACAACAGTAAAAGGGACAG 1039
Qy  241  gaccagatggtagaatcaccaagaagaaatgatatcgaactcttttgtgcctagtaaggttgtc 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1040 GACCAGATGGTAGAATCACCAAGAAGAAATGATATCGAACTCTTTTGTGCCTAGTAAAGTTGCTC 1099
Qy  301  ctgctccggcagctgtttgtgtcctccacaagtccctggaatggcaccagttcctacacggtg 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1100 CTCTCCGGCAGCTGTTTGTGTCCTCCCACAGGTCCTGGAATGGCACCAGTTCCTACACGGTG 1159
Qy  361  tcttcacagatatcccaatcagcaacattcgtcgggttattgcaacagcaggattaatgcaat 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1160 TCTTCACAGATATCCCAATCAGCAACATTCGTCGGGTTATTGCAACAGCAGGATTAATGCAAT 1219
Qv  421  caaagcaaaccataccatcattattaccttctctatcgtaaatatgggagaagtttgt 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1220 CAAAGCAAAACCATACCTCATTAATTACCTTCTCTATCGTAAATATGGGAGAAGTTTGT 1279
Qy  481  tggtacggaaagaaactataatagatattagaggagaagcaaaaatttctgtcaaatgact 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1280 TGGTACGGAAAGAAAGAACTTAATAAGATATTAGAGGAGAAGCAAAAATTTCTGTCAAATGACT 1339
Qy  541  tcatcatcataaaagtcttcagcttggcatgtttaaaagttccccgaagcaaatcttcttga 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1340 TCATCATAAAAAGCTTCAACTTTGGCATGTTTAAAAGTTCCCGAAGCAAATTCTTCTTGA 1399
Qy  601  tggacacagttataagacaaatcatgtgttgtgatgtcggtcagtactcctg 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1400 TGGACACAGTTATAAGACAAAATCATGTTGTTGATGTCGGTCAGTACTCCTG 1459
Qy  661  caggactcaatcaccattattgtgtttaatgtcacatatataaaagaggagtgaaaccattgcta 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1460 CAGGACTCAATCACCATTATTGTGTTTAATGTCACATATATAAAAGGAGTGGAAAACCATTGCTA 1519
Qy  721  atgatgttgttctttagcaacacaagcaagagaaggtaaactacacgccacatgaattcc 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1520 ATGATGTTGTTCTTTACCAGCACAAGCAAGAGAGGGTAAACTACACAGCCACATGAAATTCC 1579
Qy  781  agggtgcacttttacgatctccaatttagggatgtgttggaattaagaatttctctgcta 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1580 AGGGTGGCACTTTTACGATCTCCAATTTAGGGATGTGTTGGAATTAAGAATTTCTCTGCTA 1639
Qy  841  ttattaacccacctcaagcaatgtgttattttctgctcagaggaataaactggtcc 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1640 TTATTAACCCACCTCAAGCATGTGTTATTTTTGGCTTCAGAGGATAAAACTGGTCC 1699
Qy  901  ctgcagataatgaaaagggttggctagcataagtgtctgttaccactcagttgtg 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1700 CTGCAGATAATGAAAAAGGGTTGGCTAGCATGAGTGTCTGTTACACTCAGTTGTG 1759
Qy  961  atcaccgggtgtggtggatggagcagtcgagtggagcccaagtggctt 1001
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1760 ATCACCGGGTGGTGGTGGATGGAGCAGTTGGAGCCCAAGTGGCTT 1800
```

```
Query Match             100.0%;  Score 1001;  DB 9;  Length 2696;
Best Local Similarity   100.0%;  Pred. No. 2e-234;
Matches 1001;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1  atatatcagcatttgtgactataggccaaccgaagtaacagattaaaaccacaagtgc 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1555  ATATATCAGCATTTGTGACTATAGGCCAACCGAAGTAACAGATTAAAAACCACAAGTGC 1614
Qy   61  caccacctaccccaccccgggtggccgctgttcctcccaactccccagccttagctcca 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1615  CACCACCTACCCCACCCCCGGGTGGCCGCTGTTCCTCCCAACTCCCCAGCCTTAGCTCCTA 1674
Qy  121  cacctttcagcgcaccctgcccagctactcctgctgtggaccaaagggaagggtgttgttagcc 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1675  CACCTTTCAGCACCCTGCCCAGCTACTCCTGCTGTGGACCAAAGGGAAGGGTGTTGTTAGCC 1734
Qy  181  ctcttgcaaagaagttggcagtagagaaaggggatctcttaccacaagtaaaagggacag 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1735  CTCTTGCAAAGAAGTTGGCAGTAGAGAAAGGGGATTGATCTTACACAACAGTAAAAGGGACAG 1794
Qy  241  gaccagatggtagaatcaccaagaagaaatcgactcttttgtgcctagtaaggttgctc 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1795  GACCAGATGGTAGAATCACCAAGAAGAAATCGACTCTTTTGTGCCTAGTAAAGTTGCTC 1854
Qy  301  ctgtctccggcagctgtttgtgtcctccacaggtccctacacggttcctacacggtg 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1855  CTGCTCCGGCAGCTGTTTGTGTCCTCCCACAGGTCCTGGAATGGCACCAGTTCCTACACGGTG 1914
Qy  361  tcttcacagatatcccaatcagcaacattcgtcgggttattgcaacagcaggattaatgcaat 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1915  TCTTCACAGATATCCCAATCAGCAACATTCGTCGGGTTATTGCAACAGCAGGATTAATGCAAT 1974
Qy  421  caaagcaaaccataccatcattattacctttctctatcgtaaatatgggagaagtttgt 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1975  CAAAGCAAAACCATACCTCATTAATTACCTTCTCTATCGTAAATATGGGAGAAGTTTGT 2034
```

```
RESULT   4
LOCUS       HUMIGMBC                 2540 bp    mRNA    linear   PRI 15-FEB-1995
DEFINITION  HUMIGMBC
            Homo sapiens M2 mitochondrial autoantigen dihydrolipoamide
            acetyltransferase mRNA, complete cds.
ACCESSION   J03866
VERSION     J03866.1  GI:619443
KEYWORDS    .
SOURCE      M2 mitochondrial autoantigen; autoantigen.
            Homo sapiens (individual isolate patient with primary biliary
            cirrhosis) placenta cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2540)
  AUTHORS   Coppel,R.L., McNeilage,L.J., Surh,C.D., Van de Water,J.,
            Spithill,T.W., Whittingham,S. and Gershwin,M.E.
  TITLE     Primary structure of the human M2 mitochondrial autoantigen of
            primary biliary cirrhosis: dihydrolipoamide acetyltransferase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (19), 7317-7321 (1988)
  MEDLINE   89017189
COMMENT     On Jan 6, 1995 this sequence version replaced gi:340510.
FEATURES             Location/Qualifiers
     source          1..2540
                     /organism="Homo sapiens"
                     /isolate="patient with primary biliary cirrhosis"
                     /db_xref="taxon:9606"
                     /tissue_type="placenta"
     CDS             636..2477
                     /EC_number="2.3.1.12"
                     /codon_start=1
                     /product="dihydrolipoamide acetyltransferase"
                     /protein_id="AAA62253.1"
                     /db_xref="GI:619444"
                     /translation="MSPHCSTTYLRIGIRTTMFWKTTEGRDGKMAVQEFSEFGLLLQL
                     LGSPGRRYSLPHQKVPLPSLSPTMQAGTIARWEKKEGDKINEGDLIAEVETDKATV"
ORIGIN
BASE COUNT      666 a    644 c    603 g    627 t
```

```
Query Match          96.0%;  Score 961;  DB 9;  Length 2540;
Best Local Similarity 98.2%;  Pred. No. 1.2e-224;
Matches  983;  Conservative  0;  Mismatches  15;  Indels  3;  Gaps  1;
```

```
Qy      1 atatatcagcattgctgactataggccaaacacgaagtaacaagatttaaaaccaagtgc 60
          ||||||||||||||||||||||||||||||||| ||||||||| ||||||||||||||||
Db   1432 ATATATCAGCATTGCTGACTATAGGCCAAACACGAAGTAACAAGATTTAAAACCACAAGTGC 1491

Qy     61 caccaacctaccccacccccgttggccgcgttcctcccaactccccagccctttagctcta 120
          ||||||||||||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db   1492 CACCAACCTACCCCACCCCCGGTGGCCGCGTTCCTCCCAACTCCCCAGCCCTTTAGCTCCTA 1551

Qy    121 cacctttcagccacctgcccagctactcctgctggaacccaaaagggaagctttggtagcc 180
          ||||| ||||||||||||||||| |||||||||||||||||||||||||||| ||||||||
Db   1552 CACCTTTCGACACCCTGCCCAGCTACTCCTGCTGGAACCCAAAGGGAAGGGTGTTTTGGTAGCC 1611

Qy    181 ctcttgcaaagaagaggtttgcagttagaagaaaggagaagttgatcttacacaagtggacag 240
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1612 CTCTTGCAAAGAAGAGGTTTGGCAGTAGAGAAAGGGAGTTGATCTTACACAAGTAAAAGGACAG 1671

Qy    241 gaccagatggtagaatcaccaagagaagagatatcgactctcttttgtcccagtaagttgctc 300
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1672 GACCAGATGGTAGAATCACCAAGAGAAGAGATATCGACTCTCTTTTGTGCCCAGTAAAGTTGCTC 1731

Qy    301 ctgtccccggcagcgctgtgtgcctcccacaggtcctggaatggcaccagtcctccacagtg 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1732 CTGCTCCCGGCAGCGCTGTGTGCCTCCCACAGGTCCTGGAATGGCACCAGTCCTCCACAGTG 1791

Qy    361 tcttcacagatatcccaatcagccaacattcgtcgggtgtattgcacagcgattaatgcaat 420
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1792 TCTTCACAGATATCCCAATCAGCCAACATTCGTCGGGTGTATTGCACAGCGATTAATGCAAT 1851

Qy    421 caaagcaaacctatacctcattaccttctatcatcgataaatatgggagaagtttgt 480
          |||||||||||||||||||||||||||||| |  |||||||||||||||||||||||
Db   1852 CAAAGCAAACCATACCTCATTAATTACCTTCTATCGTGTAAATAT---GGAGAAGTTTTGT 1908

Qy    481 tggtacggaagagaacttaatagaatattagaagggagagaagcaaaattctgtcaatgact 540
          |||||||||||||||||||||||||||||| ||||||||||||| |||||||||||||||||
Db   1909 TGGTACGGAAAGGAGAACTTAATAGAATATTAGAAGGGAGAAGCAAAATTTCGTGTCAATGACT 1968

Qy    541 tcatcataaaagcttcagcttggcatgtttaaaagttcccgaagcaaatcttcttcttgga 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1969 TCATCATAAAAGCTTCAGCTTGGCATGTTTAAAAGTTCCCGAAGCAAATCTTCTTCTTGGA 2028

Qy    601 tggacacagttataagacaaaatcatgttgttgatgtcagttgtcggtcagtactcctg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2029 TGGACACAGTTATAAGACAAAATCATGTTGTTGATGTCAGTTGTCGGTCAGTACTCCTG 2088

Qy    661 caggactcatcacacctattgtgtttaatgcacatataaaagggagtggaaaccattgcta 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2089 CAGGACTCATCACACCTATTGTGTTTAATGCACATATAAAAGGGAGTGGAAAACCATTGCTA 2148

Qy    721 atgatgttgttcctttagcaacacaaagcaagagagggtaaactacagcccatgaattcc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2149 ATGATGTTGTTCCTTTAGCAACACAAAGCAAGAGAGGGTAAACTACAGCCCACATGAATTCC 2208

Qy    781 agggtggcaactttttacgatctccaattaggaatgttttgattaagaaattcctctgcta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2209 AGGGTGGCACTTTTTACGATCTCCAATTAGGAATGTTTTGATTAAGAAATTCCTCTGCTA 2268
```

```
Qy    481 tggtacggaaagaacttaatagaatattagaagggagaggaagcaaaaattctgtcaatgact 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2035 TGGTACGGAAAGAACTTAATAGAATATTAGAAGGGAGAGGAGAAGCAAAAATTCTGTCAATGACT 2094

Qy    541 tcatcataaaagcttcagcttggcatgtttaaaagttcccgaagcaaatcttcttcttgga 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2095 TCATCATAAAAGCTTCAGCTTGGCATGTTTAAAAGTTCCCGAAGCAAATCTTCTTCTTGGA 2154

Qy    601 tggacacagttataagacaaaatcatgttgttgatgtcagttgtcggtcagtactcctg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2155 TGGACACAGTTATAAGACAAAATCATGTTGTTGATGTCAGTTGTCGGTCAGTACTCCTG 2214

Qy    661 caggactcatcacacctattgtgtttaatgcacatataaaagggagtggaaaccattgcta 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2215 CAGGACTCATCACACCTATTGTGTTTAATGCACATATAAAAGGGAGTGGAAAACCATTGCTA 2274

Qy    721 atgatgttgttcctttagcaacacaaagcaagagagggtaaactacagcccatgaattcc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2275 ATGATGTTGTTCCTTTAGCAACACAAAGCAAGAGAGGGTAAACTACAGCCCACATGAATTCC 2334

Qy    781 agggtggcaactttttacgatctccaattaggaatgttttgattaagaaattcctctgcta 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2335 AGGGTGGCACTTTTTACGATCTCCAATTAGGAATGTTTTGATTAAGAAATTCCTCTGCTA 2394

Qy    841 ttattaacccactccaagcacgtgtatttggcaattggtgctcagggatataaactggtcc 900
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2395 TTATTAACCCACTCCAAGCACGTGTATTTGGCAATTGGTGCTCAGGGATAAAACTGGTCC 2454

Qy    901 ctgcagataatgaaaagggtttgatgtggctagcatgatgtctgttacacactcaagtgtg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2455 CTGCAGATAATGAAAAGGGTTTGATGTGGCTAGCATGATGTCTGTTACACACTCAAGTGTG 2514

Qy    961 ataccacggtggtggatggagagtggagcccagtggtcctt 1001
          ||||||||||||||||||||||||||||||||||||||||||
Db   2515 ATCACCGGGTGGTGGATGGAGAGTGGAGCCCAGTGGCTT 2555
```

```
GFESLRECYMAKILVAEGTRDVPIGAIICIIVGKPEDIEAFKNYILDSSAAPTPQAAP
APPPAATASPPPTSAQAPGSSSYPPHMQVLLPALSPTMMGTVQRWEKRVGEKLSEGDL
LAEIETDKATIGFEVQEEGYLAKILLVPEGGTRDVPLGTPLCIIVEKEADISAFADYRPT
EVTDLKPQVPPPPPVAAVPPQPLAPTPSTPCPATPAGPKGRVFVDPLAKKLAVE
KGIDLTQVKGTGPDGRITKKDIDSFVPSKVAPAPAAVVPPTGPGMAPVPTGVFTDIPI
SNIRRVIAQRLMQSKQIIPHYLLSCKYGEVLLVRKELNKILEGRSKISVNDFIIKAS
ALACLKVPEANSSWMDIVIRQNHIVDVSVAVSTPAGLITPIVFNAHIKGVETIANDVV
SLATKAREGKLQPHEFQGGTFTISNLGMFGIKNFSAIINPPQACILAIGASEDKLVPA
DNEKGFDVASMSVTLSCDHRVVDGAVGAQWLAEFRKYLEKPITMLL"
```

```
Qy    841 ttattaacccaccccaaggoatgtgtattggcgattggtctcttcaagagaggggata-aactggtcc 900
          |||||||||||| ||||| |||||| ||| |||| |||| ||||||||||||| |  |||||||
Db   2269 TTATTAACCCACCTCAAGCCATGTATTTTGGCAATTGGTGTCCTTCAGAGATAAACTGGTCC 2328

Qy    901 ctgcagataatgaaaaagggtttgatgtggctagacatgatgtctgttacactcagttgtg 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2329 CTGCAGATAATGAAAAAGGGTTTGATGTGGCTAGACATGATGTCTGTTACACTCAGTTGTG 2388

Qy    961 atcaccggggtggtggatggggagcagttggagcccagtggctt 1001
          |||||||||||||||||||||||||||||||||||||||||||
Db   2389 ATCACCGGGGTGGTGGATGGGGAGCAGTTGGAGCCCAGTGGCTT 2429
```

RESULT 5

LOCUS   AB036739   2543 bp   mRNA   linear   MAM 03-JUL-2001
DEFINITION Sus scrofa mRNA for dihydrolipoamide acetyltransferase, complete
            cds.
ACCESSION   AB036739
VERSION   AB036739.1 GI:14587785
KEYWORDS   Dihydrolipoamide acetyltransferase; dihydrolipoamide
            acetyltransferase.
SOURCE   Sus scrofa
  ORGANISM Sus scrofa cDNA to mRNA.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 2543)
  AUTHORS   Koike,K.
  TITLE   Isolation and characterization of the cDNA encoding the porcine
            dihydrolipoamide acetyltransferase component of the
            pyruvate dehydrogenase complex
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2543)
  AUTHORS   Koike,K.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-JAN-2000) Kichiko Koike, Nagasaki University School
            of Medicine, Department of Pathological Biochemistry; Sakamoto
            1-12-4, Nagasaki, Nagasaki 852-8523, Japan
            (E-mail:Kichiko.Koike@mb5.seikyou.ne.jp, Tel:81-95-849-7098,
            Fax:81-95-883-5065)
FEATURES   Location/Qualifiers
     source   1..2543
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
     CDS   6..1949
                     /EC_number="2.3.1.12"
                     /codon_start=1
                     /product="dihydrolipoamide acetyltransferase"
                     /protein_id="BAB61720.1"
                     /db_xref="GI:14587786"
                     /translation="MWRVCAPRAQNAPRAGPEGPCGARWTALREEPGAPCAAPRAGSVPAR
                     CSSTTRGYGRSRALCGWSATRGATPQNRILLQLWGSPNRRWYSLPPHQKVPLSLSPT
                     MQAGTIARWEKKEGDKINEGELIVAEVETDKATVGFESLEECYMAKILVAEGRDVPVG
                     AICITVEKPEDIEAFKNYTLDSSAAPAPTPAAAPAPTPSAQAPGSSYPTH
                     MQVVLPALSPTMTMGTVQRWEKKVGEKLSEGDLLAEIETDKATIGFEVQEEGYLAKIL
                     IPEGTRDVPLGTPLCIIVEKEADIPAFADYRPTEVTDLKPQPPPPSPVTPVPPAPQ
                     PVAPTPAATRPAPAPGKGRLFVSPLAKKLASEKGIDLTQIKGTGPDGRIIKEDISF
                     VPKAAPPPAAAVPPPSPGVAPVPTGVFTDIPISNIRRVIAQRLMQSKQTIPHYLSV
                     DVNMGEVLLVRKELNKMLEGRSKISVNDFIIKASALACLKVPEANSSWLDTVIRQNHV
                     VDISVAVSTPAGLITPIVFNAHIKGLEIIANDVVSLATKAREGKLQPHEVQGGTFTIS
                     NLGMFGIKNFSAINPQACILAVGASEDRLFPADNEKGFDVASMMSVTLSCDHRVVD
                     GAVGAQWLAEFRKYLEKPITMLL"
     sig_peptide   6..263
     mat_peptide   264..1946
                     /product="dihydrolipoamide acetyltransferase"
     polyA_signal   2145..2150
BASE COUNT   724 a   563 c   576 g   680 t
ORIGIN

Query Match   87.1%; Score 871.4; DB 4; Length 2543;
Best Local Similarity 91.9%; Pred. No. 1.1e-202;
Matches 920; Conservative   0; Mismatches 81; Indels   0; Gaps   0;
```

```
Qy      1 atatatcaagcatttgctgactataggccaacgtaagctaagctaagtaacacacaagtgc 60
          |||| ||| |||||| |||||||||  || ||||   || ||||||  ||||| |||||
Db    901 ATATACACGACATTTGTGACTATCGCGACCAACTGAAGTAACTGAGTTAAAACCACAGAC 960

Qy     61 caccaacttaccccacccccccgtggccgcgtgttcctcccaagtaacctcccccagtggctc 120
          |||||||| ||| |||||||||||| ||||||||||||||| |||  ||||||| |||||
Db    961 CACCAACTACCCGTCCCCGGGTGACCCCGTGTTCCTCCAGCTCCCAAACCCTA 1020

Qy    121 cacctttcagtcagacccctgcccaagctactcctgctggaccaaaaaagggtgttgttagc 180
          ||||| || | | | ||||| ||||||||||||||||| |||||| | ||| ||||||| |||
Db   1021 CACCTGCACCCGCCCCAGCTACTCCTGCCGGACCAAAGGGAAGGGTTGTTTGTTAGCC 1080

Qy    181 ctcttgcaaagaaagttggcagtagagaaaaggagttgatcttacacaagtaaaagggacag 240
          ||||||| ||||||| |||||| ||||||||||| | ||| |  ||||  |||| |||||
Db   1081 CTCTTGCCAAAGAAAATTGGCATCAGAGAAAAGGCCATTCGCCATCCGCCAACTGAAGTAACCACAGAC 1140

Qy    241 gaccagagatggtagaatcaaccagagaggagatcgactctcttttgccttagtctc 300
          |||||| |||| ||||||||| |||||||| || |||  || || ||| ||| |||||
Db   1141 GACCAGAATGGCAGAATCAATCAAGAAGGACAATGAGCTGATCCTGGCTGAAGTTGAGACTGAC 1200

Qy    301 ctgtccggcagctgttgtgcctccacaagtgtcctggaatggcaccagttcctcacagtgt 360
          ||||| | |||| ||||||||||||||||| | |||| |||| |||||||| || ||| |
Db   1201 CTACTCCAGCAGCTGTGTTCCTCCCCACCAGTGGTGGGCACCAGTTCCCAGGGG 1260

Qy    361 tcttcacagtatcccaatcagcaacattcgtcgggttcattggcccagcgattaatgcaat 420
          ||||||| || |||| | |||| |||  | |  |||  ||| || ||| ||| |||||
Db   1261 TCTTCACAGATATCCCAATCAGCAACATTCGTCGAGTTATTGCACAGCGGTTAATGCAAT 1320

Qy    421 caaagcaaaaccataacctcattaccttctatcatgtaaatatggaaagtttgt 480
          ||||||||| |||| |||  |||  |||| |||| ||||| ||||| |||| |||
Db   1321 TCATTATAAAAGCTTCAGCCTGGCATGTTAAAAAGTTCCTGAACCAAATCTCTTTGC 1380

Qy    481 tggtacggagaagacttataagatattagagagagcaaaatttctgtcaatgact 540
          |||| |||||| |||||| |||||||| |||| | ||||||||||||||| | |
Db   1381 TGGGACAGGTTATAAGACAAAAATCATGTTTGGAGGTGCGGCAAAATTTCGTGCAATGATT 1440

Qy    541 tcatcataaaagcttcagctttgcatgtttaaaaagttcccgaagcaaacatttgt 600
          || | |||| ||| ||| ||| |||| ||||| |||||| || ||| | |  |||
Db   1441 TCATTATAAAAGCTTCAGCCTTGGCATGTTAAAAAGTTGTTTTGATATTATCATTGC 1500

Qy    601 caggactcacacctatgtgttcaatcatgttgttgatgtcggtcaagtaactcctg 660
          ||||| || ||||| |||| || || |||| |||| || | ||| ||| || |||
Db   1501 CAGGACGTTATAAGACAAAATCATGTTTGGAGGTGCGGCAAAATTTCGTGCAATGATT 1560

Qy    661 caggactcacactatgtgttcaatcgaacaatataaaaggagagactggaaaccttgcta 720
          |||||| || ||| ||||| ||| | || || |||| |||||||||| | ||| |||||
Db   1561 CAGGACTCATCACACCTATGTGTTATTAATGCACATATAAAAGGACTGGAAACCATTGCTA 1620

Qy    721 atgatgttgttcttagcaacaccaagagcaaggggtaaactacagccacatgaattcc 780
          ||||| ||||| ||| |||  |||| |||   |||  | | |  |||||||| | |||
Db   1621 ATGATGTTGTTTCTTTAGCAACCAAAGCACGAGGGTAAACTACAGCCACATGAGGTCC 1680

Qy    781 agggtggcacttttacgatctccaattaggaaatgtttggaataaagtttcctgtgcta 840
          |||||||| || ||| ||| ||  || ||| |  |||||||||||| |||||||| ||
Db   1681 AGGGTGGCACTTTTACAAATCTGGAATTAGGAAAGTTTGGAATTAAGAGACTTCTGCTA 1740

Qy    841 ttattaacccacccaagcatgtgtattttggctagacatggtgctcagaggatcggtcc 900
          |||||||||| ||  ||| ||| ||| |||||||| | || || |||| |||  | |||
Db   1741 TTATTAACCCTCCTCAAGCATGTGGCTTCAGAGGATGAGCTGATCCTGGCTTCAGACTGTTC 1800

Qy    901 ctgcagataatgaaaaagggtttgacgtagctagactagatgagctcgttacactcagttgtg 960
          |||||| |||| |||  ||||| | ||| ||||||||| |||  ||| ||||| ||||||||
Db   1801 CAGCCGGATAATGAAAAAGGATTTGACGTAGCTAGACTAGATGANGTCTCGTTACACTCAGCTGTG 1860

Qy    961 atcaccggggtggtggatggggagcagttggagcccagtggctt 1001
          |||||||||||||||||||| ||||||||||||||||||||||
Db   1861 ATCATCGGGGTTGTTGGATGGGGAGCAGTTGGAGCCCAGTGGCTT 1901
```

RESULT 6

AY042265

LOCUS   AY042265   1700 bp   mRNA   linear   ROD 01-NOV-2001

```
DEFINITION  Mus musculus dihydrolipoamide S-acetyltransferase precursor, mRNA,
            partial cds.
ACCESSION   AY044265
VERSION     AY044265.1  GI:16580127
KEYWORDS    .
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1700)
AUTHORS     Wang,L., Kaneko,S., Kagaya,M., Ohno,H., Honda,M. and Kobayashi,K.
TITLE       Molecular cloning, characterization and expression of
            dihydrolipoamide acetyltransferase component of murine pyruvate
            dehydrogenase complex in NIH3T3 cell
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 1700)
AUTHORS     Wang,L., Kaneko,S., Kagaya,M., Ohno,H., Honda,M. and Kobayashi,K.
TITLE       Direct Submission
JOURNAL     Submitted (09-JUL-2001) First Department of Internal Medicine,
            Kanazawa University School of Medicine, Takara Machi 13-1,
            Kanazawa, Ishkawa 920-8640, Japan
.TURES      Location/Qualifiers
source      1..1700
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /cell_line="BML1-ME"
            /tissue_type="hepatoma"
CDS         <1..>1680
            /note="PDC-E2; component of the pyruvate dehydrogenase
            complex"
            /codon_start=1
            /product="dihydrolipoamide S-acetyltransferase precursor"
            /protein_id="AAL02400.1"
            /db_xref="GI:16580128"
            /translation="YYSLPPHQKVPLPSLSPTMQAGTIARWEKKEGEKISEGDLIAEV
            ETDKATVGFESLEECVNAKILVPEGTRDVPVGSIICIIVEKPQDIEAFKNYTLDLAAS
            AAPQAAPAAPAAPAAPAAPAAPSASAPGSSYPTHMQIVLPAPSPTMTMGTVQRWEKKVGE
            KLSEGDLLAEIETDKAIIGFEVQEEGYLAKIXVPEGTRDVPLGAPXIIIVEKOEDIAA
            FADYRPTEVTSIKPQAAPAAPPVAANVPPTPQPVAPPTPSAAPAGPKGRVFVSPLAKKL
            AAEKGIDLTQVKGTGPECRIIKKDIDSFVPSKAAPAAAAMAPPGPRVAPAPAGVFTD
            IPISNIRVIAORLMOSKOTIPHYYLSVDVNMGEVLLVRKELNKMLEGKGKISVNDFI
            SDVVSLASKARECKLQPHEFQGGTFTISNLGMFGIKNFSAIINPPQACILAIGASEDK
            LIPADNEKGFDVASVSMSVTLSCDHRVVDGAVGAQWLAEFKKYLEKPITMLL"
mat_peptide ?..1677
            /product="dihydrolipoamide S-acetyltransferase"
BASE COUNT   459 a   414 c   423 g   402 t
ORIGIN
```

```
Query Match    76.8%;   Score 768.6;  DB 10;  Length 1700;
Best Local Similarity    86.4%;   Pred. No. 1.5e-177;
Matches  865;  Conservative   0;  Mismatches 124;  Indels  12;  Gaps  1;

Qy    1 atatatcagcattgtgactataggccaaccgaagtaacagatattaaaaccaaagtgc 60
         |||||| ||||| ||||| ||||||||||||| ||| ||||| |||| ||||| |||||
Db  644 ATATAGCAGCATTGCAGACTACAGGCCAACACAGAAGTGACCACCAGCTTAAAGCCACAG 703

Qy   61 caccacctaccccaccccagtggccgcgtgggccgcgtgttcctcccaagccttagctcta 120
         |||||||||||||||| |||||||||||||| |||| ||||||||||||| ||||||| ||
Db  704 CACCACCTACCCCCCCCGGTGGCCAGCTGTTCCTCCCACTCCCCAGCCTTAGCACCTA 763

Qy  121 cacccttcagcacctgcccagctcactcctgctggaccaaagggaaggtgttgttagcc 180
         |||| |            ||||||| |||||||||||||||||||||||| ||| |||| |
Db  764 CCCC-------------CTCAGCTGCTCCTGCTCGGACCAAAGGGAAGGTTGTTCGTTAGTC 811

Qy  181 ctcttgcaaagaagttggcagtagagaagaggattgatcttacacacaagaggacag 240
         ||||||||| |||||||||||| ||||| ||| |||| || ||||  ||||||||||
Db  812 CTCTTGCCGAAGAAGTTGGCAGCAGAGAAAGGGAGCGTGAGAGGGACGG 871

Qy  241 gaccagagtgggtagaataacagaagatatcgaactcttttgtccctagtaaagttgtc 300
         |||||||| |||  |||||||| || ||  | ||||||||||||||  || ||| ||| |
Db  872 GACCAGAGAGGCAGGAATCATCAAGAAAGACAATTGACTCTTTTGTCCCTTCTAAGGCTGCT 931

Qy  301 ctgctccggcagctgttgtgcctcccacaggtcctggaatggagccagttcctacaggtg 360
         ||||| |||||| ||| ||| |||||||||| ||| |  ||||| ||||| ||| |||||
Db  932 CTGCTGCCAGCAGCGTGCTGCCTCCCACAGGCTCCCCCGGGTTCCCCCGGGTCCAAGAGCTG 991

Qy  361 tcttcaacgagatatcccaatcagcaacaattcgtcggggttattggcacagcggatataatgcat 420
         |||||| ||| ||| ||||||||||||||||||||||| ||| |||||||||| |||
Db  992 TCTTCACAGATACCCCATCAGCAACAATTCGTCGGAGTGATTGGCACAGCAGT 1051

Qy  421 caaagcaaaccatacctcattaccttctatcgtatgtaaatatgggaagaagtttgt 480
         | |||||  |||| ||||| ||||  ||||  ||||| ||||||||| || |||| ||
Db 1052 CGAAGCAGACTATACCTCATTATTACCTTTCTCGTCGGATAAATATGGGAGGAAGTTGCTGT 1111

Qy  481 tggtacggaagaaacttaataagatattagaaggagaagcaaaaattctctgtcaatgact 540
         ||||||||||| |||||||||||||| |||||||||| |||||||||||||| || |||||
Db 1112 TGGTACGGAAGGAAGGAACTTAATAAGATGCTTGAAGGGAAAAGGGAAAAATCTCGTGACACT 1171

Qy  541 tcatcataaagcttcaagctttggcatgtttaaaagtccccgaagcaaacaaatttcttcttgga 600
         |||| | ||||||||| |||||||| ||||| |||||||| |||||||||   |||||| ||||
Db 1172 TCATTATTAAAGGCTTCAGCTTTGGCCGTGTTCTGAAAGTTCCCGAAGCAAACTCGTCTTGGA 1231

Qy  601 atgatgttgtttcttctttagcaaccaaagcaagcagaagagggtaaaactacccacaccacatcc 660
         | |||| |||| ||| |||||| ||||||||||  ||| |||||||||||||| |   || ||||
Db 1232 TGGACACAGTTATTAAGGCAAAAATCATGTGGTTGACGTCAGTGTTGCTGTCAGTAGCCCCTG 1291

Qy  661 caggactcatcacacctattgtgttctttaatgcacatataaaagggaagtggaaaccacttgcta 720
         |||||| |||||||||||||||||| |||||| ||||||||||||| ||||| ||| ||| ||||
Db 1292 CAGGACTTATCACCCCTATTGTGTTCAACGACACACATAAAGAGACTTGGAAAACATTGCTA 1351

Qy  721 atgatgttgtttcttagcaaccaaagcaagaagaggggtaaactacagccaacatgaatcc 780
         ||||| |||| |||| |||| |||||| || ||||||||||||| ||||| ||| |||||||
Db 1352 GTGATGTGTTGTTTCTTTAGCCTCCAAAGGCGAGAGAGAGGGTAAAACTTCAGCCTCATGAGTTCC 1411

Qy  781 aggggtggcacttttacgatctcccaattaggaatgttctggaattaagagaatttctgcta 840
         ||||||||||| |||| | ||||| |||||||| |||| |||||||||||||||||| || |
Db 1412 AGGGTGGCAACATTTACAATCTCCAACTTAGGGATGTTTGGAATTAAGAGAATTTCTCTGCAA 1471

Qy  841 ttattaacccacctccaagcatgtatttggcaattgtgcttcaagaggaataaactggtcc 900
         |||||||  ||||||| |||||| || | | ||||||||||||||| ||||| |||| ||||
Db 1472 TTATTAATCCACCTCAGGCATGCATTGTAATCTGCAACCGGTCGCTTCAAGGGATAAAACTGTGATCC 1531

Qy  901 ctgcagataatgaaaaagggtttgatgtggctagcagtgtctgttcacctcaagttgtg 960
         | |||||||||||| ||||| ||||||||| ||| || |||||| ||| ||||| ||||
Db 1532 CAGCAGATAATGAACAAGGCCTTTGATGTGTGGCCTAGTGTCTGTTACACTCAGCTGTG 1591

Qy  961 atcaccggttggtatggagtcagctgggacccaagtggctt 1001
         |||| | | ||| |||||||||| |||||| ||||| |||||
Db 1592 ATCATCGAGTTGTGGATGGAGCAGTTGGAGCCCAGTGGCTT 1632
```

```
RESULT 7
RATPDCE2
LOCUS       RATPDCE2     2266 bp    mRNA    linear  ROD 20-APR-2000
DEFINITION  Rattus norvegicus mRNA for dihydrolipoamide acetyltransferase.
ACCESSION   D10655
VERSION     D10655.1  GI:220837
KEYWORDS    dihydrolipoamide acetyltransferase.
SOURCE      Rattus norvegicus (strain:Sprague-Dawley) cDNA to mRNA,
            clone_lib:lambda gt11.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 2266)
AUTHORS     Matuda,S., Nakano,K., Ohta,S., Shimura,M., Yamanaka,T.,
            Nakagawa,S., Titani,K. and Miyata,T.
TITLE       Molecular cloning of dihydrolipoamide acetyltransferase of the rat
            pyruvate dehydrogenase complex: sequence comparison and
            evolutionary relationship to other dihydrolipoamide
            acyltransferases
JOURNAL     Biochim. Biophys. Acta 1131 (1), 114-118 (1992)
MEDLINE     92256482
COMMENT     These data kindly submitted in computer readable form by: Sadayuki
```

Matsuda
Department of Biology
Institute of
Fitness and Sports
Shiromizucho, Kanoya
Kagoshima 891-23
Japan
    Phone: 0994-46-4111
    Fax: 0994-46-2831.

Kanoya National

FEATURES   Location/Qualifiers
    source   1..2266
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone_lib="lambda gt11"
    gene   1..1626
                     /gene="PCD-E2"
    CDS   <1..1626
                     /gene="PCD-E2"
                     /codon_start=1
                     /product="dihydrolipoamide acetyltransferase"
                     /protein_id="BAA01504.1"
                     /db_xref="GI:220838"
                     /translation="PTMQAGTIARWEKKEGEKISEGDLIAEVETDKATVGFESLEECY
                     MAKILVPEGRDVPIGCICICITIVEKPEQDIEAFKNYTLDSATAATQAPAPAPAAAP
                     AAPSASAPGGSSYPVHMQIVLPALSPTMTMGTVQRWEKKVGEKLSEGDLLAEIETDKAT
                     IGFEVQEEGYLAKILVPEGTRDVPLGTPLCIIVEKEQDIAAFADYRPTEVTSLRPQAP
                     PVPPVAAVPPIQPLAPTPSAAPAGPKGRVFVSPLAKKLAAEKGIDLTQVKGTGPEG
                     GRIIKKDIDSFVPTKAAPAAAAAPPGPRVAPTPAGVFIDIPISNIRRVIAQRLMQSK
                     QTIPHYLSVDVNDGEVLLVRKELNKMLEGKGIKVSVNDFIIKASALACLKVPEANSSW
                     MDTVIRQNHVHVDVSVAVSTPAGLIITPIVFNAHIKGLETIASDVVSLASKAREGKLQPH
                     EFQGGTFTISNLGMFGIKNFSAIINPQACLIAIGASEDKLIPADNEKGFDVASVMSV
                     TLSCDHRVVDGAVGAQWLAEFKKYLEKPVIMLL"

BASE COUNT   654 a   492 c   512 g   608 t
ORIGIN

Query Match   74.8%; Score 747.8; DB 10; Length 2266;
Best Local Similarity   85.8%; Pred. No. 1.9e-172;
Matches   859; Conservative   0; Mismatches 127; Indels 15; Gaps 2;

Qy   1 atatatcaagcatttgtcgactataggccaaccgaagtaaacagattaaaacaacaagtgc   60
             |||||||||||||||||||||||||||||||||||| ||||| |||| |||||| || |
Db   593 AATATAGACAGTCAATTTGCAGACTACAGGCCAACACAGAAGTCGACCAAGCTTAAAGCCACAGGCAC   652

Qy   61 caccaacctaccccaccccacccgttggccgcgtgttcctccaaactccccagccttagctccta   120
             ||||||||| ||||||||||||||||||||| |||||||||||||||||||||||||| |||
Db   653 CACCAACCTGTCTCCACCCCCACGTGGCACGTGTTCCTCCCCATCCCCAGCCTTTAGCACCTA   712

Qy   121 cacttcagcaccctgcccagctactcctgtggaccaaaggaaggagagtgtgttggtaagcc   180
             |||||||| || |||||||||||||||| ||||||||||||||||||||| ||| |||||||
Db   713 CTCAGCCGCTCCTGCTGGACCAAAGGGAAAGGGGTGTTCGTTAGCC   760

Qy   181 ctcttgcaaagaagttggcagtagagagagaggattgatctctacacagcagtaaaagggacag   240
             |||||||||||| ||||||||||||||||| |||| ||||||||||||| |||| ||||||| |
Db   761 CTCTTGCAAAGAAAATTGGCGACAGCAGAGAAAGGGATTGACCTCACCACCAAGTTAAAGGGACGG   820

Qy   241 gaccagatggtagaatcaccaagagaaagatatcgaactctttgtgcctagtaagttgtc   300
             |||||||||||||||| ||| ||||||||||||||||||||||||||||||| |||||||
Db   821 GACCAGAGGCAGAAATCATCAACAAGGACAATTGCACTCTTTGTGCCTACTAAGGCTGCTC   880

Qy   301 ctgctccggcagctgttgtgcctccaccacaggtcgaaccagttcctacacaggtg   360
             |||| ||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   881 CTGCGCGTGCAGCAGCTG----CTCCCCCCGGGTCCAAGAGTGGCGACCAACTCCTGCGCAGGTG   937

Qy   361 tcttcacaagatatcccaatcagcaacattcgtcgggttattgcaacgacgattaatgcaat   420
             ||||| |||||||||||||||||||| ||||||| ||||||||||||| ||| |||||||
Db   938 TCTTCATAGACATCCCCATCAGCAACATTCGTCGAGTTATCGCACAGCGCCTCATGCAGT   997

Qy   421 caaagcaaaccatacctcacttcattacctttctatcgatgtaaatatggggagaagtttttgt   480
             ||||||||||| ||| |||||| |||||||||| ||| ||||| ||| |||||||||||||
Db   998 CGAAGCAGACTATACCTCATTATTACCTTTCTGTTGATGTAAATATGGGGAGAAGTGCTGT   1057

BASE COUNT   368 a   348 c   335 g   319 t
ORIGIN

RESULT   8
A05167
LOCUS   A05167   1370 bp   DNA   linear   PAT 04-MAY-1993
DEFINITION   Synthetic pBMIT nucleotide sequence of the 70kd mitochondrial
             antigen of PBC.
ACCESSION   A05167
VERSION   A05167.1   GI:345027
KEYWORDS   .
SOURCE   synthetic construct.
  ORGANISM   synthetic construct
             artificial sequence.
REFERENCE   1 (bases 1 to 1370)
  AUTHORS   .
  TITLE   PRIMARY BILIARY CIRRHOSIS AUTOANTIGEN
  JOURNAL   Patent: WO 8804689-A 2 30-JUN-1988;
FEATURES   Location/Qualifiers
    source   1..1370
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
    CDS   <1..>1370
                     /codon_start=1
                     /transl_table=11
                     /product="70kd mitochondrial antigen of PBC"
                     /protein_id="CAA00430.1"
                     /db_xref="GI:345028"
                     /translation="GPEARKNYILDSATAATQAAPAPAAAPAAPSASAPGGSSYP
                     VHMQIVLPALSPTMTMGTVQRWEKKVGEKLSEGDLLAEIETDKATIGFEVQEEGYLAK
                     ILVPEGTRDVPLGTPLCIIVEKEQDIAAFADYRPTEVTSLRPQAPPVPPVAAVPPI
                     QPLAPTPSAAPAGPKGRVFVSPLAKKLAAEKGIDLTQVKGTGPEGRIIKKDIDSFVP
                     TKAAPAAAAAPPGPRVAPTPAGVFIDIPISNIRRVIAQRLMQSKQTIPHYLSVDVN
                     MGEVLLVRKELNKMLEGGKGIKVSVNDFIIKASALACLKVPEANSSWMDTVIRQNHVVDV
                     SVAVSTPAGLIITPIVFNAHIKGLETIASDVVSLASKAREGKLQPHEFQGGTFTISNLG
                     MFGIKNFSAIINPQACLIAIGASEDKLIPADNEKGFDVASVMSVTHSAVIIELMMEQ
                     LEPSGLL"

Qy   481 tggtacggaaagaacttaaagtatattagaaagggagaagcaaaatttctgtcaatgact   540
             |||||||||||||||||||||||||| ||||||||||||||| ||||||| || ||||||
Db   1058 TGGTGCGGAAGGAGAACTTAATAGATGCTGTTGAAGGTAAAGGAAGAAAATCTCCGTCAATGAACT   1117

Qy   541 tcatccataaaagcttcagcttggcatgttaaaagttccgaagcaaatctcttctttggga   600
             ||||||||| |||||||||||| |||||||| |||| ||||| |||||| ||| |||| ||
Db   1178 TCATCATAAAAGCCTTCAGCTTTGGCCTGCCTGCCCTGAAGCTTCCTGAAGCAAACGTCATCTTGGA   1177

Qy   601 tggacacaagttataagacaaaaatcatgttgttgatgtcagtcggtcagtactcctg   660
             ||||||||||| || ||||| ||||| ||||||||||| ||| ||||| | ||||||||
Db   1178 TGGACACAAGTTATTACGACAAAAATCATCGTGGTTGATGTCAGTCGTTGCGTGTCAGTACCCTG   1237

Qy   661 caggaccatcaccacctattgtgttaatgtcacatataaaaggagagtgggaacctgcta   720
             |||||| |||| |||||||||||||| |||||||||||| ||| |||| |||||||||||
Db   1238 CAGGACTTATCACCCTATTGTGTTTAATGCACACATATAAAAGGACACAGGAAAACCATTGCTA   1297

Qy   721 atgatgttgttcttttagcaaccaaagccaagagagggtaaactacagcccacatgaattcc   780
             |||||||||||| ||||||||||||||| ||||| |||| || || |||||| |||||||
Db   1298 GTGATGTTGTTCTTTTGTTTACCCTCCAAAGCCAAGAGAGGGTAAACTTCACCCTCACGAGTTCC   1357

Qy   781 agggtggcacttttacgatctccaatttaggaatgtttggaattaagaaattctctgcta   840
             |||| |||||||||||||| ||||||||||||||| ||||| |||||||||||||||||
Db   1358 AGGGTGGGCACTTTTACGAATCTCCAACTTAGGCAATGTTTCGGAATTAAGAAATTTCTCTGCGA   1417

Qy   841 ttattaaccaacctcaagcacatgtatcttttggcaattggtgctcaaggagataaactggtcc   900
             ||||||||||||||||||||||||||| |||| ||||| ||||||||||| |||| |||||||
Db   1418 TTATTAACCACCTCAAGCACATGTATTTTTGGGCAATGGGTCGGCATGTATTTTGGCAATTGGTGCTCAAGGAGATAAACTGGATCC   1477

Qy   901 ctgcaagataatgagaaaagggttcagctagcatgatgatgtctcttgttacactcaagttgtg   960
             |||||||||||||||||||||||| ||||||||||||||| |||| ||||||| |||||||||
Db   1478 CTGCAGATAATGAGAAAAGGCTTTGACGTGGCGTAGTGTCAATGTCTGTTACACACTCAGCTGTG   1537

Qy   961 ataccgggtgttgggatgggagcagcccaagtgcttt   1001
             |||| |||||||||||||||||||||||||||||||||
Db   1538 ATCATCGAGTTGTGGGATGGGAGCAGTTGGAGCCCCAGTGGCTT   1578

BASE COUNT   368 a   348 c   335 g   319 t
ORIGIN

```
Query Match   73.5%;   Score 735.8;   DB 6;   Length 1370;
Best Local Similarity   85.6%;   Pred. No. 1.5e-169;
Matches 859;   Conservative   0;   Mismatches 127;   Indels 17;   Gaps 3;

Qy    1 atatatcaagcatttgtgactataggccaaccgaagtaacagattaaaaccacaagtgc   60
        |||| ||| |||||| ||| |||| ||||||||| ||||||| ||||||| || ||||
Db  380 ATATAGCAGCATTTGCAGACTACAGGCCAACAGAAGTGACCAGCTTAAAGCCACAGGCAC  439

Qy   61 caccacctacccaccacccccggttggccgctgttcctccaactccccagcttagctccta  120
        |||||||| | |||||||| | |||| || ||||||||||| | ||||||| ||| |||||
Db  440 CACCACCTGTCCCACCCCCAGTGGCCAGCTGTTCCTCCCATCCCCAGCCCTTTAGCACCTA  499

Qy  121 caccttcagcaccctgcccagctactcctgtctgaccaaaggaaggggtgtttgttagcc  180
        |||          ||||| | ||||||| |||| |||||||| |||||||||| |||||||
Db  500 CCCC------------CTCAGCGCGTCCTCGTGGACCAAAAGGGAAGGGGTGTTCGTTAGCC  547

Qy  181 ctcttgcaaagaagttgggcagtagaaagggaggaaaggggattgacctcacacaagtaaaagggacag  240
        ||||||||||||| ||| |||| ||||| |||||| |||||||||||| | ||||||||| | ||||||
Db  548 CTCTTGCAAAGAAAATTGGCACGAGAGAAAAGGGATTGACCTCACCAAAGGTTCGTTAGCC          607

Qy  241 gaccagatggtagaatcaccaagaaagatatcgacctcttttgtgcctagtaaagttgctc  300
Db  608 GACCAGAAGCAGAAATCAAGAACAGGACACATTGACTCTTTTGCCTACTAAGGCTCGTC     667

Qy  301 ctgctccggcagctgttgtgctcccaccaagtgtcctcggaattggcaccagttcctacaggtg  360
Db  668 CTGCCGCTGCCAGCAGCTG---CTCCCCGGGGTCCAAGAGTGGCACCAACTCCTGCAGGTG     724

Qy  361 tcttccagatatcccaatcagcaacaattcgtcgggttattgcacagcgattaatgcaat  420
Db  725 TCTTCAAGACTATACCTCATTATTACCTTCTCATCGATGTAAAATATGGGAGAGGTGCTGT  784

Qy  421 caaagcaaaccataccttcattattacctttctatcgatgtaaatatgggagagagttttgt  480
Db  785 CGAAGCAGACTATACCTCATTATTACCTTCTCATCGATGTAAAATATGGGAGAGGTGCTGT  844

Qy  481 tggttacggaaagaacttaataagatattagagaaggggagacaaaaattctgtcaatgact  540
Db  845 TGGTGCGGAAGGAAACTTAATAAGATGCTTGAAGGTAGAAGGAAAAATCTCCGTCAATGACT  904

Qy  541 tcatcatcaaaagcttcagcttggcatgttaaagttcccggaagcaaagttcccttgga  600
Db  905 TCATCATAAAAAGCCTTCAGCTTTGGCCTGCCCTGAAAGTTCGAAGCAAAACTCATCTTGGA  964

Qy  601 tggacacagttatagagacaaaatcatgtgttgtcagttcagtgtgcggtcagtactcctg  660
Db  965 TGGACACAGTTATACGACAAAAATCATGTGGTTCATGTCAGCGTGCTTCAGTCACCCCTG   1024

Qy  661 caggactcaatcacaccttattgtgttaatgcaacatatataaaaaggaagtggaaaccaattgcta  720
Db  965 TGGACACAGTTATACGACAAAAATCATGTGGTTCATGTCAGCGTGCTTCAGTCACCCCTA   1024

Qy  721 atgatgttgttttacggatctccagcaaccaagcaaagctaaactacagccacatgaattcc  780
Db 1025 CAGGACTTATCACCCGTATTGTGTTTAATGCACACATAAAAGGACTGGAAACCATTGCTA   1084

Qy  781 agggtggcacttttacgatctccaattagattaaggaattaggaattttctctgcta  840
Db 1085 GTGATGTTGTTTCTTTAGCCAGCAAAGCAAGAGAGGGTAAAACTTCAGCCACATCAGCCT  1144

Qy  841 ttattaaccacctcaagacatgtattttggcaattagggatgacctggtcc  900
Db 1145 AGGGTGGGACATTACAATCCCAACTTAGGGATGTTCGGAATTAAGAATTTCTCGCGA   1204

Qy  901 ctgcagataatgcaaaaagggtttgatggctaagcaatgatgtctgt--tacactcagttg  958
Db 1205 TTATTAACCCACTCAGCGCATGTTATTTTGGCAATTGGTGGCTTCCGAGGATAAACTGATCC  1264

Qy  959 tgatcaccgggttgggtggatgggagcagttggagcccagtggctt  1001
Db 1265 CTGCAGATAATGAGAAAGGCTTTGACGTGGCTAGTGTGTGTCACACTCAGCTG   1324

Qy  181 ctcttgcaaagaagttggcagtagaaagggaggaaaggggattgacctcacacaagtaaaagggacag  240
Db 1325 TGATCATCGAGTTGTGGATGGAGCAGTTGGACCCCAGTGGCTT   1367
```

```
Qy   241 gaccagatgtagaatcaccaagaagaagatatcgactcttttgtgctctagtaaagtgctc 300
         ||||||  || || ||||||| ||||||   || |||||||||||| | ||| ||  |||||
Db   608 GACCAGAAGGCAGAAATCATCAAGAAGACAATTGACTCTTTTGTGCCTACTAAGGCTGTC 667

Qy   301 ctgctccggcagctgttgtcctccccaaggtcctggaatggcaccagttcctacaggtg 360
Db   668 CTGCCCGGGGCTGCACAGCTG---CTCCCCCGGGGTCCAAGACAGTGGCCAACTCCTGCACGTG 724

Qy   361 tcttcacagatatccaatcagcaacacattcgtcgggttattgcacacgcgattaatgcaat 420
Db   725 TCTTCATAGACACATCCCCATCAGCAACATTCGTCGAGTTGACTGCTTGCACAGT 784

Qy   421 caaagcaaaccatacctcattattacctttcatcgatgtaaatatggagaaagttttgt 480
Db   785 CGAAGCAGACTATACCTCATTATTACCTTTCTGTGATGTAAATATGGGAGAGGTGCTGT 844

Qy   481 tggtacggaaagaaacttaataagatattagaagggagaagcaaaatttctgtcaaatgact 540
Db   845 TGGTGCCGAAGAAACTTAATAAGATGCTTGAAGGAAAAAATCTCCGTCAATGACT 904

Qy   541 tcatcataaaagcttcagctttggcatgttaaaagttcccgaagcaaatttcttcttgga 600
Db   905 TCATCATAAAAGCTTCAGCTTTGGCCTGCTGAAAGTTCCTGAAGCAAATTCAATCTTGGA 964

Qy   601 tggacacagttataagacaaaaatcatgttgttgatgtcagtgtcggtcagtactcctg 660
Db   965 TGGACACAGTTATACGACAAAATCATGTGGTTGATGTCAGCGGTTGCTGAAGAAGGTGCTGT 1024

Qy   661 caggactcatcacacctattgtgttcaatgcacatataaaaggagtggaaaccattgtca 720
Db  1025 CAGGACTTATCACCCCTATTGTGTTCAATGCACATATAAAAGGACTGGAAACCATTGCTA 1084

Qy   721 atgtagttgttctttagcaacaacaagcaagagaggtaaactacagcacacatgaactcc 780
Db  1085 GTGATGTTGTTTCTTTAGCCTCCAAAGCAAGAGAGGGTAAACTTCAGCCTCACGAGTTCC 1144

Qy   781 agggtggcactttacgatcctccaaatttaggaatgttgattagacaattctctctgcta 840
Db  1145 AGGGTGGGACATTTACAATCTCCAACTTAGGGAATGTTGGGAATAAGAATTTCTCTGCA 1204

Qy   841 ttattaacccacctcaagcaatgtattcttttggcaattggtcgctcaagaggataatggtcc 900
Db  1205 TTATTAACCCACCTCAAGCCATGTATTTTTGGCAATTGGTCGCTCAAGAGGATAAGTGCTC 1264

Qy   901 ctgcagatcaatgaaaagggtttgattggctagcatgatgtctgt---tacactcagttg 958
Db  1265 CTCGAGTAATGACAAAAGGCTTTGACGTGGCTAGTGTCGCGTCACACACTCAGCTG 1324

Qy   959 tgatcaccgggtcggtgatgggagcgcagttggagcccagtggctt 1001
Db  1325 TGATCATCGAGTTGTGGGATGGAGCGAGTTGGAGCCCAGTGGCTT 1367
```

RESULT 10
```
LOCUS       RATPBCAA        1370 bp    mRNA    linear   ROD 27-APR-1993
DEFINITION  Rat primary biliary cirrhosis mitochondrial autoantigen mRNA.
ACCESSION   M16075
VERSION     M16075.1  GI:206037
KEYWORDS    autoantigen; mitochondrial antigen.
SOURCE      Rat liver, cDNA to mRNA, clone pRMIT.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1370)
AUTHORS     Gershwin,M.E., Mackay,I.R., Sturgess,A. and Coppel,R.L.
TITLE       Identification and specificity of a cDNA encoding the 70 kd
            mitochondrial antigen recognized in primary biliary cirrhosis
JOURNAL     J. Immunol. 138, 3525-3531 (1987)
MEDLINE     87196380
COMMENT     Clean copy of sequence [1] kindly provided by M.E.Gershwin
```

```
FEATURES             Location/Qualifiers
     source          (15-JUN-1987);
                     1..1370
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     CDS             <1..>1370
                     /note="primary biliary cirrhosis autoantigen"
                     /codon_start=1
                     /protein_id="AAA41813.1"
                     /db_xref="GI:206038"
                     /translation="GPEAFKNYILDSAFATAQNAPAPARAPAAPAPARPSASARGSSYP
                     VHMQIVLPALSPPIMTMGTVQRWEKKVGEKLSEGDLLAEIETHKATIGFEVQEBGYLAK
                     ILVPEGTRDVPLGTPLCIIVEKQEDIAAFADYRPTEVTSLKFQAPPVPPVAAAVPPI
                     PQPLAPTPSAAPAGPKGRVFVSPLAKKLAAEEKGIDLTQVKGTPGRIIKKDIDSFVP
                     TKAAPAAAAPPGPRVAPTPAGVFIDIPISNTRRVIAQRLMQSKQTIPHYLSVDVN
                     MGEVLLVRRELNKMLEGGKISVNDFIIKASALACLKVPEANSSWMDTVIRQNHRVDV
                     SVAVSTPAGLITPIVFNAHIKGLETIASDVVSLASKAREGKLQPHEFQGGTFTISNLG
                     MPGIKNFSAIINPQACLLAIGASEDKLIPADNEKGFDVASVMSVTHSAVIIELKMEQ
                     LEPSGLL"
BASE COUNT      368 a    348 c    335 g    319 t
ORIGIN      Unreported.
```

```
Query Match   73.5%;  Score 735.8;  DB 10;  Length 1370;
Best Local Similarity   85.6%;
Matches  859;  Conservative   0;  Mismatches 127;  Indels  17;  Gaps  3;
                                   Pred. No.1.5e-169;
```

```
Qy     1 atatatcagcacatttgctgactataggccaaccgaagtaaccagattaaaaccacaagtgc 60
Db   380 ATATAGCAGCATTTGCAGACTACACGGCCAACAGAGAAGTGACCAGGCTTAAAGGCCAC 439

Qy    61 caccacctacccaccaccccgggtgcgcctgttctccaactccccagcgccctttagctcccta 120
Db   440 CACCAACCTGTCCCACCCACCCAGTGACGAGTGTTCCCCCACCACCCCACGGGCCAGCCTA 499

Qy   121 caccttcagcaccctgcccagctactcccgtgtgaccaaagggaaggggtgtttgttagcc 180
Db   500 CCCC----------CTCAGCGCGCTCCTGCCTGGACCAAAGGCGAAGGGGTGTTAGCC 547

Qy   181 ctcttgcaaagaagttggcagtagagaaagggattggatcttacacaagtaaaagggacag 240
Db   548 CTCTTGCAAAGAAGTTGGCAGTGGAGAAATGGGATTGGATCTTACAACAAGTAAAAGGGACGG 607

Qy   241 gaccagatgtagaatcaccaagaagaagatatcgactcttttgtgctctagtaaagtgctc 300
Db   608 GACCAGAAGGCAGAAATCATCAAGAAGACAATTGACTCTTTTGTGCCTACTAAGGCTGTC 667

Qy   301 ctgctccggcagctgttgtcctccccaaggtcctggaatggcaccagttcctacaggtg 360
Db   668 CTGCGCGGGGCTGCACAGCTG---CTCCCCCGGGGTCCAAGACAGTGGCCAACTCCTGCACGTG 724

Qy   361 tcttcacagatatccaatcagcaacacattcgtcgggttattgcacacgcgattaatgcaat 420
Db   725 TCTTCATAGACACATCCCCATCAGCAACATTCGTCGAGTTGACTGCTTGCACAGT 784

Qy   421 caaagcaaaccatacctcattattacctttcatcgatgtaaatatggagaaagttttgt 480
Db   785 CGAAGCAGACTATACCTCATTATTACCTTTCTGTGATGTAAATATGGGAGAGGTGCTGT 844

Qy   481 tggtacggaaagaaacttaataagatattagaagggagaagcaaaatttctgtcaaatgact 540
Db   845 TGGTGCCGAAGAAACTTAATAAGATGCTTGAAGGAAAAAATCTCCGTCAATGACT 904

Qy   541 tcatcataaaagcttcagcttggcatgttaaaagttcccgaagcaaatttcttcttgga 600
Db   905 TCATCATAAAAGCTTCAGCTTTGGCCTGCTGAAAGTTCCTGAAGCAAATTCAATCTTGGA 964

Qy   601 tggacacagttataagacaaaaatcatgttgttgatgtcagtgtcggtcagtactcctg 660
Db   965 TGGACACAGTTATACGACAAAATCATGTGGTTGATGTCAGCGGTTGCTGAAGAAGGTGCTGT 1024

Qy   661 caggactcatcacacctattgtgttcaatgcacatataaaaggagtggaaaccattgtca 720
Db   965 CAGGACTTATCACCCCTATTGTGTTCAATGCACATATAAAAGGACTGGAAACCATTGCTA 720
```

RESULT 11

```
LOCUS       A05169                   966 bp    DNA     linear   PAT 04-MAY-1993
DEFINITION  Synthetic pRMIT 2.2kb cDNA insert of the 70kd mitochondrial antigen
            of PBC.
ACCESSION   A05169
VERSION     A05169.1  GI:345029
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 966)
  AUTHORS
  TITLE     PRIMARY BILIARY CIRRHOSIS AUTOANTIGEN
  JOURNAL   Patent: WO 8804689-A 4 30-JUN-1988;
FEATURES             Location/Qualifiers
     source          1..966
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
     CDS             <1..>966
                     /codon_start=1
                     /trans_table=11
                     /product="Human 70kd mitochondrial antigen of PBC"
                     /protein_id="CAA00431.1"
                     /db_xref="GI:345030"
                     /translation="PGSSYPPHMQVLLPALSPTMTMGTVQRWHKKVGEKLSEGDLLAE
                     IETDKATIGFEVQEEGYLAKILVPEGTRDVPLGTPLCIIVEKEADISAFADYRPTEVT
                     DLKPQVPPPTPPVAAVPTPQPLAPPSAPCPATPAGPKGRVFVSPLAKKLAVEKGI
                     DLTQVKGTGPDGRITKKDIDSEVPSKVAPAPAAVPPTGPGMAPVPTGVFTDIPISNI
                     RRVIAQRLMQSKQTIPHYYLSIDVNMGEVLLVRKELNKILEGRSKISVNDFIIKASAL
                     ACLKVPEANSSWMDTVIRQNHVVDVSVAVSTPAGLIPIVFNAHIK"
ORIGIN
BASE COUNT    289 a    229 c    208 g    240 t
```

```
Query Match          70.0%;  Score 701;  DB 6;  Length 966;
Best Local Similarity  100.0%;
Matches  701;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 atatatcagcattgctgactataggccaaccgaagtaacagatttaaaaccacaagtgc 60
         ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   266 ATATATCAGCATTTGCTGACTATAGGCCAACCGAAGTAACAGATTTAAAACCACAAGTGC 325

Qy    61 caccacctaccccacccccGgtggccGctgttcctccaactccccagcccttagctcca 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   326 CACCACCTACCCCACCCCCCGGTGGCCGCTGTTCCTCCAACTCCCCAGCCCTTTAGCTCCA 385

Qy   121 caccttcagcaccctgccccagctactcctgctgtggaccaaaggaaggtgttgttagcc 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   386 CACCTTCAGCACCCTGCCCCAGCTACTCCTGCTGTGGACCAAAGGAAGGGTGTTGTTAGCC 445

Qy   721 atgatgttgttcttcttagtcaaccaaaagcaaagagaggaggggtaaactacagccacatgaattcc 780
         |||||||||||| ||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  1025 CAGGACTTATCACACCCTATTGTTGTTTAATGCACACATAAAAGGACTGGAAACCATTGCTA 1084

Db  1085 GTGATGTTGTTGTTTCTTAGCCCTCCAAAAGCAAAGAGAGGGGTAAACTTCAGCCTCACGAGTTCC 1144

Qy   781 agggtggcacttttacgatctccaattagtgaaatgtttgggaattaggagaattcctctcgctca 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1145 AGGGTGGGACATTTACAATCTCCAACTTAGGGATGTTCCAACTTAAGGAATTTCTCTCGGA 1204

Qy   841 ttattaacccaccctcaagcatgtattttggccaattggtgcttcagaggagataaactggtcc 900
         |||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  1205 TTATTAACCCACCCTCAGGCATGTATTTTTGGCCAATTGGTGCTTCCGAGGATAAAACTGATCC 1264

Qy   901 ctgcagaataatgcaaagagtttgatgtggctagcatgatgatgtctgt--tacactcagttg 958
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  1265 CTGCAGAATAATGCAAAGAGTTTGACGTGGCTAGCATGTCTGTGCACACTCAGCTG 1324

Qy   959 tgatcaccgggtggtggatggaggaagcagttggagacccaagtggctt 1001
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1325 TGATCATCGAGTTGTGTGGATGGAGCAGTTGGAGCCCCAGTGGCTT 1367
```

RESULT 12

```
LOCUS       HSDHACTYL                660 bp    mRNA    linear   PRI 10-JAN-1994
DEFINITION  H.sapiens mRNA for dihydrolipoamide S-acetyltransferase.
ACCESSION   X13822
VERSION     X13822.1  GI:30523
KEYWORDS    acetyltransferase; dihydrolipoamide S-acetyltransferase; lipoate
            acetyltransferase; pyruvate dehydrogenase.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Moehario,L.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-1988) Moehario L.H., Department of Biochemistry
            and Centre for Molecular Biology and Medicine, Monash University,
            Clayton, Victoria 3168, Australia
REFERENCE   2  (bases 1 to 660)
  AUTHORS   Moehario,L.H., Smooker,P.M., Devenish,R.J., Mackay,I.R.,
            Gershwin,M.E. and Marzuki,S.
  TITLE     Nucleotide sequence of a cDNA encoding the lipoate acetyl
            transferase (E2) of human heart pyruvate dehydrogenase complex
            differs from that of human placenta
  JOURNAL   Biochem. Int. 20 (2), 417-422 (1990)
  MEDLINE   90197673
COMMENT     Related sequence J03866.
FEATURES             Location/Qualifiers
     source          1..660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="heart"
                     /clone_lib="cDNA library in lambda gt11"
     CDS             <1..>660
                     /EC_number="2.3.1.12"

Qy   181 ccttgcaaagaagttggcagttagagagaaagggggattgatcttacacaagtaaaagggacag 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   446 CTCTTGCAAAGAAGTTGGCAGTAGAGAAGGATTGATCTTACACAAGTAAAAGGGACAG 505

Qy   241 gaccagatggtagaatcaccaggagaagggatatcgaactcttctggtcctagtaaagttgctc 300
         |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   506 GACCAGATGGTAGAATCACCAAGAAGGATATCGACTCTTTTGGTCCTAGTAAAGTTGCTC 565

Qy   301 ctgctccggcaggctgttgtgtgtccccacaggtccctggaatggcaccagttcctacaggtg 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   566 CTGCTCCGGCAGGCTGTTGTGTGTCCCCACAGGTCCCTGGAATGGCACCAGTTCCTACAGGTG 625

Qy   361 tcttcacagatatcccaatcagcaacattcgtcgggttattggcacagcggattaatgcaat 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   626 TCTTCACAGATATCCCAATCAGCAACATTCGTCGGGTTATTGCCACAGCGGATTAAATGCAAT 685

Qy   421 caaagcaaaaccatcacccctcaattattacccttcctctatcgtgttaaatatgggagacgttt 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   686 CAAAGCAAAACCATCACCCCTCATTATTACCCTTTCTATCGTGTTAAATATGGGGAGAAGTTT 745

Qy   481 tggtacggaaagaacttaataagatattagaagggagaagcaaaatttctgtcaaatgact 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   746 TGGTACGGAAAGAACTTAATAAGATATTAGAGGGAGAAGCAAAATTTCTGTCAAATGACT 805

Qy   541 tcattcataaagcttcagctttggccatgttttaaaagtccgaacgaagcaaaatttcttcttgga 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   806 TCATCATAAAGCTTCAGCTTTGGCCATGTTTTAAAAGTCCGAACGAAGCAAAATTCTTCTTGGA 865

Qy   601 tggacaacagttataagacaaaacaatgtttgttgatgtgttgcggtcagttgcggtcgtactcctg 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   866 TGGACACAGTTATAAGACAAAAATCATGTTTGTTGATGTGTTGCGGTGTTCGGGTGTACTCCTG 925

Qy   661 caggagactcatcacacctattgttgttaatgcacatataaa 701
         ||||||||||||||||||||||||||||||||||||||||||||
Db   926 CAGGACTCATCACACCTATTGTGTTTAATGCACATATAAA 966
```

```
                    /codon_start=1
                    /product="dihydrolipoamide S-acetyltransferase"
                    /protein_id="CAA32052.1"
                    /db_xref="GI:30524"
                    /db_xref="SPTREMBL:Q01991"
                    /translation="PPTPSAPCPATPAGPKGRVFVSPLAKKLAVEKGIDLTQVKGTGP
                    DGRIYKDIDSFVPSKVAPAPAAVPPTPGQMAPVTPGVFTDIPISNIRRVIAQRLMQ
                    SKQTIPHYLSIDVNMGEVLLVRKELNKILEGRSKISVNDFIIKRSALACLKVPEANS
                    SWMDTVIRQNHVDVSVAVSTPAGLITPIVFNAHIKGVETIANDVVSLATKAREGKLQ
                    PH"
BASE COUNT      200 a     140 c     143 g     177 t
ORIGIN
```

```
Query Match       65.5%;     Score 655.8;   DB 9;    Length 660;
Best Local Similarity  99.7%;     Pred. No. 5.4e-150;
Matches  657;   Conservative  0;   Mismatches  2;   Indels  0;   Gaps  0;
```

```
Ov    115 ctcctcaacctttcagcacctgcaaaagaaagttgggcagtagagagagaaaggagggtgttg 174
          ||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      2 CTCCTACACCTTCAGCACCTGCAAAAGAAAGTTGGGCAGTAGAGAGAGAAAGGAGGGTGTTTG 61

Qy    175 ttagccctcttgcacaagatgttggtcagtagagagagaaaggagggtgtttg          234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TTAGCCCTCTTGCACAAGAAGTTGGCAGTAGAGAGAGAAAGGAGGGTGTTTG           121

Qy    235 ggacaggaccagatgtagaatcaccaagagaggagaagatacgacctctttgtgtctagtaaag 294
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 GGACAGGACCAGATGGTAGAAATCACCAAGAGAGGAGAATATGCGACCTCTTTGTGCCTAGTAAAG 181

Qy    295 ttgctcctgctccggcagctgttggtcctccacacagagttcctggaatggcaccagttccta  354
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 TTGCTCCTGCTCCGGCAGCTGTTGGTCCTCCTGGAATGGCACCAGTTCCTA            241

Qy    355 caggtgtcttcacagatatcccaatccagcaacatcgtcgggttattgtcacagcgattaa    414
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CAGGTGTCTTCACAGATATCCCAATCCAGCAACATCGTCGGGTTATTGTCACAGCGATTAA    301

Qy    415 tgcaatacaaagcaaacatcatcatctcatcttcatcgatgtaaaatagggagag         474
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TGCAATACAAAGCAAACCATTACCTTCTCATCCATATTACCTTTCTCATCGATGTAAAAATAGGGAGAAG 361

Qy    475 tttgtgttcggtaacggaaagaaacttaataagatattagaaggggagagaagcaaaattctgtca  534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 TTTTGTGTTCGGTACGGAAAGAACACTTAATAAGATATTAGAAGGGGAGAAGCAAAATTTCTGTCA 421

Qy    535 atgacttcatcataaaagcttcagcttggcatgttcaagttcccgaagcaaaattctt       594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ATGACTTCATCATAAAAAGCTTCAGCTTTGGCATGTTTAAAAGTTTCCCAAGCAAAATTCTT 481

Qy    595 cttggatggacacagttataagacaaaatcatgttgtgatgtcaagttcggtcagta        654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 CTTGGATGGACACAGTTATAAGACAAAATCATGTTGTTGATGTCAGTGTTGCGGTCAGTA 541

Qy    655 ctcctgcaggactcatcacactcctattgtgttaatgcacatataaaagtggaaacca       714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 CTCCTGCAGGACTCACTCACACCTTATTGTGTTAATGCACATATAAAAGGAGTGGAAACCA 601

Qy    715 ttgctaatgatgttgttgtttcttttagtcaaccaaaagcaaaggaaggtagaactacaagaccacat 773
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 TTGCTAATGATGTTGTTTCTTTTAGCAACCAAAAGCAAAGGAAGGTAGAACTACAGCCACAT 660
```

```
RESULT  13
BC003202
LOCUS       BC003202     2038 bp    mRNA    linear   ROD 12-JUL-2001
DEFINITION  Mus musculus, clone IMAGE:3586777, mRNA, partial cds.
ACCESSION   BC003202
VERSION     BC003202.1  GI:13096817
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2038)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 10 Row: g Column: 8
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1..2038
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3586777"
                     /tissue_type="Mammary tumor metastatized to lung. Tumor
                     arose spontaneously from a senescent normal mammary
                     (clonal) outgrowth infected with the virus MMTV."
                     /clone_lib="NCI_CGAP_Lu29"
                     /lab_host="DH10B"
                     /vector: pCMV-SPORT6"
     CDS             <1..371
                     /codon_start=3
                     /product="Unknown (protein for IMAGE:3586777)"
                     /protein_id="AAH03202.1"
                     /db_xref="GI:13096818"
                     /translation="HASAHIKGLETIASDVVSLASKAREGKLQPHEFQGGTFTTSNLG
                     MFGIKNFSAIINPQACILAIGASEDKLIPADNEKGFDVVSVWSVTLSCDHRVVDGAV
                     GAQWLAEFRKYLEKPITMLL"
BASE COUNT      632 a     369 c     419 g     618 t
ORIGIN
```

```
Query Match       26.4%;     Score 264;     DB 10;    Length 2038;
Best Local Similarity  90.4%;     Pred. No. 4.9e-54;
Matches  282;   Conservative  0;   Mismatches  30;   Indels  0;   Gaps  0;
```

```
Qy    690 gcacatataaaagtggaaaccattgctaatgatgttgttgtttcttttagtcaaccaaaagca 749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 GCACACATAAAAAGGACTGGAAAACCATTGCTAGTGATGTTGTTGTTTCTTTAGCCTCAAAAGCG 71

Qy    750 agagaggggtaaactacaagcccactggaattccaaggtggcactttacgatcctccaattta 809
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     72 AGAGAGGGTAAACTTCAGCCTCATGAGTTCCAGGGTGGGAACATTTACAACTCCAACTTA 131

Qy    810 ggaatgtttcggaattaagaaatttctctgtattattaaccaaccccaagcatgtatttg 869
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    132 GGGATGTTTGGAATTAAGAAATTTTCTGCAATTATTAACCCACCTGCAATGTGATTCTG 191

Qy    870 gcaattggtgcttcaggagataaactggtccctgcagataatgaaaaagggtttgatgtg 929
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 GCAATCGGTGCTTCAGGAGATAAACTGATCCCAGCAGATAATGAGAAAGGCTTTGATGTG 251

Qy    930 gctagcatgtcatgtctgttacctcagttgatcaccggggtggtggatggagagcagttgga 989
```

```
Db    252 GTTAGTTGTGATGTCCGTTACACTCAGCTGTGATCATCCAGTTGTGTGGATGGAGCAGTTGGA 311
          ||||  |||||||| |||||||||| |||||||| |  |||||||||| || ||||||||||||||
Qy    990 gcccagtggctt 1001
          ||||||||||||
Db    312 GCCCAGTGGCTT 323
```

Db   REFERENCE 14
LOCUS   AX305911   319 bp   DNA   linear   PAT 11-DEC-2001
DEFINITION Sequence 662 from Patent WO0188188.
ACCESSION   AX305911
VERSION   AX305911.1 GI:17645279
KEYWORDS
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
  AUTHORS   Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
  TITLE   Method for examining ischemic conditions
  JOURNAL   Patent: WO 0188188-A 662 22-NOV-2001;
            School Juridical Person Nihon University (JP)
FEATURES   Location/Qualifiers
     source   1..319
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT   89 a   60 c   77 g   93 t
ORIGIN

```
Qy    731 ttctttagcaaccaagcaaagagagaggggtaaactacacgcaccatgaattccaggtggcac 790
          |||||||||| ||||||| |||| ||| |  |||||| |||  ||||||||||||||||||| ||
Db      1 TTCTTTAGCTCTCCAAACCGAGAGAGGGTAAACTTCAGCCTCATGAGTTCCAGGGTGGAAC 60
Qy    791 tttacgatctccaatttaggagaatgttggaattaagaaattctctgctattattaaccc 850
          |||| ||||||||  |||||||||  |||||||| ||||||| |||||| || ||||| ||
Db     61 ATTTACAAATCTCCAACTTAGGGATAGTGTTCGGAATTAAGAAATTTCTCTGCAAATTAGTTAGTAGTCC 120
Qy    851 acctcaagcaatgtattttggcaattggtgcttcaagaggaatgataaactggtccccgtgcagatga 910
          |||||||||||||| ||||||| || ||| |||||||| ||  |||| |||| |||| |||||| ||||
Db    121 ACCTCAGGCATGTATTTTCTGGCAATCGGTGCTTCATAGGATAAAACTGATCCCAGCAGATAA 180
Qy    911 tgaaaaagggtttgatgtggctagacatgaatgtctgttacactcaagttgtgatcaccggt 970
          ||||||| |||| || || || ||| ||||||||| ||||||||||||| ||||| ||||||
Db    181 TGAGAAAGGCTTTGAATGTGGCTAGTGTGTCCGTTACACTCAGCTGTGATCATCGAGT 240
Qy    971 ggtggatggagcagttggagcccagtggctt 1001
          |||||||||||||||||||||  ||||||||
Db    241 TGTGGATGGAGCAGTTGGAGCCCAGTGGCTT 271
```

RESULT 15
LOCUS   DDU06634   1833 bp   mRNA   linear   INV 08-MAR-1994
DEFINITION Dictyostelium discoideum dihydrolipoamide transacetylase mRNA,
            partial cds.
ACCESSION   U06634
VERSION   U06634.1 GI:458425
KEYWORDS
SOURCE   Dictyostelium discoideum.
  ORGANISM Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida.; Dictyostelium.
REFERENCE   1 (bases 1 to 1833)
  AUTHORS   Mueller-Taubenberger,A.
  TITLE   Dihydrolipoamide transacetylase gene from Dictyostelium discoideum
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1833)

  AUTHORS   Mueller-Taubenberger,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-FEB-1994) Annette Mueller-Taubenberger,
            Max-Planck-Institute for Biochemistry, Am Klopferspitz 18A,
            Martinsried, 82152, Germany
FEATURES   Location/Qualifiers
     source   1..1833
                     /organism="Dictyostelium discoideum"
                     /strain="AX2"
                     /db_xref="taxon:44689"
                     /clone_lib="lambda-gt11"
     CDS   <1..1779
                     /note="subunit E2 of the dehydrogenase complex; product is
                     located in the mitochondrion; putative single copy gene,
                     but 2 mRNAs of 1.5kb and 2.1kb expressed"
                     /codon_start=1
                     /product="dihydrolipoamide acetyltransferase"
                     /protein_id="AAA16511.1"
                     /db_xref="GI:458426"
                     /translation="FNNYQTKPKIFTSSNVLSFSPSSSSNVPSEILNKRSYSSKGKRI
                     TMPALSPSMTVGNIVQWKKKEGDQIKAGDVIEVETDKATMDSYEDGNGIYLAKILLPE
                     GTRGIEINKPIAIIVSKREDIESSAVKNYKPSSQASSTPVQEEAPKPQEEAPKRSTKIY
                     PAHKVGHPALSPSMETGGIASWTKKEDQIKAGDAIAEVETDRAATMDFQYEDGNGYL
                     AKILVPGGTSGTGIQINOPVCIIVKNEKDCKFADYSVEEQSSSSSSSSSSSQESPSSSSS
                     SQESTPSQSSSSQQTTRKGGERIFATPAARFEASSKGIDLSAINGTGPNRRIKADVLE
                     FVPQKQEVAQQQOQOTTTTRKPPTPTSSGEFFDIPHSNIRKVTAARLTESKQTIPHY
                     YLTMECVRDKLKLRSELNAMNVKIISVNDFIVKASLPALRDNPVVNSTWTDQFIRY
                     HNIDINAVNTPGLFTPIVRGVDMKGLNSISTSVKQLAEKAQNGKLHPSEFESGGTT
                     ISNLGMLGIKQFAAVINPPQAAILALVPQKLVSPLSNKPDSPETAILLSVTLSCDHR
                     VIDGAVGAEWLKSFKDYVENPIKLIL"
     misc_feature   103..390
                     /note="lipoyl domain 1, amino acids 35-130"
     misc_feature   244..246
                     /note="lipoyllysine residue 1, amino acid 82"
     misc_feature   472..723
                     /note="lipoyl domain 2, amino acids 158-241"
     misc_feature   610..612
                     /note="lipoyllysine residue 2, amino acid 204"
     misc_feature   889..1008
                     /note="E3-binding domain, amino acids 297-336"
     misc_feature   1078..1776
                     /note="catalytic domain, amino acids 360-592"
BASE COUNT   701 a   347 c   276 g   509 t
ORIGIN

Query Match   21.0%;   Score 210;   DB 3;   Length 1833;
Best Local Similarity   54.8%;   Pred. No. 7.9e-41;
Matches 489;   Conservative   0;   Mismatches 385;   Indels   18;   Gaps   3;

```
Qy    121 caccttcagcaccctgcccaagctactcctggtggaccaaaggaagggtgttgttagcc 180
          |||   || | |||  |||||| |||||  || ||  | |||||  ||  || ||||||
Db    845 CATCACAATCATCACACAACAACTACACACGTAAATCTGGTGAAAGAATTTCGCTACAC 904
Qy    181 ctcttgcaaagaagttggcagtagagaaaggaggatgatcttacacaagtaaaaggacag 240
          ||||   || |||||||| |||| |||  || || || ||  ||| ||| ||| ||||||
Db    905 CAGCAGCAAGAGTTGAGGCATCATCAAAAAGGAATATGATCTTTCAGCAANTCAAANGGTACTG 964
Qy    241 gacaagatggtagaatcaccaagaaggagatatcgactcttttgtgctagtaaagttgtc 300
          ||   || |||||  |||  |||||||||   |||  | ||||| |  ||| |||| || |
Db    965 GTCCAAAATAATAGAATTCTTAAAGCTGATGTTTGTTACCACACAAAAAACAANAGAAG 1024
Qy    301 ctgctccggcagctgtgtgc---------ctccacaggtcctggaatggaccagttc 351
          ||||| || ||||| ||||            |||||| ||| ||||||| | |||||
Db    1025 TTGCTCAACAATCATCACAACAACAACTACTACCACCACTAAAAAACAACTACCAACTT 1084
Qy    352 ctacaggtgtcttcacagatatcccaaatcagcaacattcgtcgggttattgcaagcgat 411
          ||    ||||| | ||||| |   || ||| ||| |||  ||||| || ||| || || |
Db    1085 CATCAGGTGAATTCACTGATATTCCACATAGTAATATTCCGTAAGGTTACTGCAGCTCGTC 1144
Qy    412 taatgcaatcaaagacaaaaccaatacctcaattattaccttctgatgtaaaaatgttatg 471
          |||| ||||| |||  ||||||    ||| || | || |||| ||||| || ||||| |||
Db    1145 TTACAGAATCAAAAACAAACCATTCCACATTATTATCTCACAATGGAATGTCGTGTTGATA 1204
```

```
Qy    472 aagtttgttggtacggaaagaacttaataagatattagaagggagaagcaaaatttctg 531
              | | ||   ||     |   || |||   |  ||||           |||||  |||||
Db   1205 AACTCTTAAAACTTCGTTCAGAATTGAATGCAAT------GAACACTGTTAAAATCTCTG 1258

Qy    532 tcaatgacttcatcataaaagcttcagcttggcatgttgtttaaaagttcccgaagcaaatt 591
              ||||| ||| || ||| | |||| | | |||||        |||  |  | |||||
Db   1259 TCAATGATTTCATTGTTAAAGCATCGCTGCCGGCCCTTCGTTGATAATCCAGTCGTAAATT 1318

Qy    592 ctttcttggatggaacaagttataagacaaaatcatgttgatgtcagtgttgcggtca 651
              | | |  | ||  ||  |    ||||  ||    |    ||| ||  || || ||
Db   1319 CAAACATGGACCGATCAAATTCATTAGAAGATATCACAACATTGATATATCAACGTTGCCGTCA 1378

Qy    652 gtactcctgcaaggactcatcacacctattgtgtttaatgcacatataaaaggagtggaaa 711
              |||     |     ||    | |      |||   | |   |      |    ||
Db   1379 ATACACCACAAGGTTTTATTCACTCCAAATCGTTAGAGGTGTCGATAATGAAAGGTCTCAACT 1438

Qy    712 ccattgctaatgatgttgttctcttagcaaccaaagcaagagaagggtaaactacagccac 771
              ||    |   |  |        |||   |||    ||   |||    |  |  ||| |
Db   1379 CTATCTCAACCTCTGTTAAACAATTGGCTCGAAAAATGGTAAAATTACACCCAT 1498

Qy    772 atgaattccagggtggcacttttacggatctccaattaggaatgtttggaattaagaatt 831
              | |  |||| |    | |      || | | | |  |  |  | ||      || |
Db   1499 CCGGAATTTGAAAGTGGTACTTTCACTATATCTCTAACTTGGGTATGCTCGGTATTAAAACAAT 1558

Qy    832 tctctgctattattaacccacctcaagcatgtgtattttggcaattggtgcttcagggata 891
              | | || ||         | | | |  |  || ||||   ||     ||  ||| |
Db   1559 TCGCACGAGTCATCAATCCACCACACAGCTGCTATCTTTGGTCGACCACAGAAACTCG 1618

Qy    892 aactgtccctgcagataa---tgaaaaaggggtttgatgtggctagacatgatgtctgtta 948
              ||  |   |  | |       |||   |  ||  ||  ||| ||     |||  |   |||
Db   1619 TGTCGTTTTTAAGCAATAAAACCAGACTCACCCATATGAAACCGCTACCCATTTTATCCGTTA 1678

Qy    949 cactcagttgtgatcacgggtgtggatggaagcagttggagcccagtggct 1000
              | ||||||| | |     | ||   | | ||  |  | |  |||    |||
Db   1679 CCTTAAGTTGTGTGATCATCCGTGTTTATCGATGGTGCTGTAGGTGCTGAATGGTT 1730
```

OM nucleic - nucleic search, using sw model

```
                         GenCore version 4.5
              Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:              August 13, 2002, 18:26:23 ; Search time 208.64 Seconds
                                                 (without alignments)
                                                 8237.305 Million cell updates/sec

Title:               US-09-761-580-1_COPY_800_1800
Perfect score:       1001
Sequence:            1 atatatcaagcattgctgac.........cagttggagccagtggctt 1001
Scoring table:       IDENTITY.NUC
                     Gapop 10.0 , Gapext 1.0
Searched:            1736436 seqs, 85857221 residues
Listing first 45 summaries

al number of hits satisfying chosen parameters:        3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%

Database :       N_Geneseq_032802:*
        1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
        2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
        3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
        4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
        5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
        6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
        7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
        8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
        9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
       10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
       11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
       12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
       13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
       14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
       15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
       16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
       17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
       18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
       19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
       20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
       21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
       21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
       22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001.DAT:*
       23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
       24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1001	100.0	2583	22	AAD07345	Human pyruvate deh
2	1000.2	99.9	3859	21	AAF221916	Human breast and o
3	735.8	73.5	1370	9	AAN82410	Plasmid pRMIT enco
4	699.4	69.9	966	9	AAN82411	Insert from cDNA p
5	250.4	25.0	360	23	AAS70149	DNA encoding novel
6	223	22.3	319	24	AB199637	Mouse ischaemic co
7	209.8	21.0	1502	23	ABL20587	Drosophila melanog
8	209.8	21.0	1850	23	ABL09105	Drosophila melanog
9	155.4	15.5	639	23	AAS70150	DNA encoding novel

```
  c   10    125.8   12.6   4145   23   ABL20586   Drosophila melanog
  c   11    125.8   12.6   4983   21   ABL09104   Drosophila melanog
  c   12    125.6   12.5   2348   21   AAC46977   Arabidopsis thalia
      13     96.5    9.7   2595   23   AAS92326   DNA encoding novel
      14     88.2    8.8  22863   21   AAA81500   Human pyruvate deh
      15     88.2    8.8 349980   21   AAF21609   Neisseria meningit
      16     88      8.8   1302   22   AAH52969   N. meningitidis pa
      17     88      8.8   3133   22   AAH54299   S. epidermidis ope
      18     88      8.8   3190   22   AAH54272   S. epidermidis gen
      19     88      8.8   3354   22   AAH54399   S. epidermidis gen
      20     88      8.8   3501   22   AAH53982   S. epidermidis gen
  c   21     85      8.5   1549   21   AAC44948   Arabidopsis thalia
  c   22     85      8.5   1708   20   AAX06839   Arabidopsis pyruva
  c   23     81.8    8.2   4858   18   AAV74690   Staphylococcus aur
  c   24     80.8    8.1   1971   23   AAS00749   DNA encoding novel
  c   25     80.4    8.0 580073   18   AAT58840   Mycoplasma genital
  c   26     79.4    7.9   1290   23   AAS52165   Staphylococcus aur
  c   27     79.4    7.9   1293   23   AAS55200   Staphylococcus aur
  c   28     78.8    7.9   1260   22   AAH53561   S. epidermidis ope
  c   29     78.8    7.9   3394   22   AAH54265   S. epidermidis gen
      30     77.8    7.8    482   21   AAC35785   S. epidermidis gen
      31     77.8    7.8   1230   23   AAS53536   zea mays DNA fragm
      32     77.6    7.8 119211   23   AAF28553   Haemophilus influe
  c   33     77.6    7.8 1830121  17   AAF42063   Genomic fragment #
  c   34     77.6    7.8   1044   23   AAS55575   Haemophilus Influe
  c   35     76      7.6  11340   19   AAV52280   Streptococcus pneu
  c   36     75.8    7.6   7953   18   AAV74463   Streptococcus pneu
      37     74.6    7.5 269223   22   AAF28554   Genomic fragment #
      38     72.4    7.2   1254   20   AAX07106   Staphylococcus aur
      39     72.4    7.2   1254   20   AAX07107   Staphylococcus aur
      40     72.4    7.2   1254   20   AAX07108   Staphylococcus aur
      41     71.8    7.2   1702   21   AAC35957   Staphylococcus aur
      42     71.8    7.2   1839   21   AAC38982   Arabidopsis thalia
      43     70.2    7.0 1038602  20   AAZ01425   Complete genome se
      44     69.8    7.0   1185   22   AAH52579   S. epidermidis ope
      45     69.8    7.0   2706   20   AAV81790   Granulocytic Ehrli
```

ALIGNMENTS

```
RESULT    1
AAD07345
ID     AAD07345 standard; DNA; 2583 BP.
XX
AC     AAD07345;
XX
DT     12-SEP-2001  (first entry)
XX
DE     Human pyruvate dehydrogenase complex E2 (PDH E2 or PDC E2) gene.
XX
Human pyruvate dehydrogenase complex E2; PDH E2; PDC E2; polymorphism;
lactic acidaemia; coronary failure; cardiac myopathy; weakness; diabetes;
muscle ataxia; peripheral vascular disease; myocardial ischaemia; asthma;
bioinformatic analysis; pharmacogenetic analysis; obesity; sepsis; drug;
genetic marker; therapy; ds.
OS     Homo sapiens.
XX
FH     Key           Location/Qualifiers
FH
FT     CDS           211..1848
FT     variation     /*tag= a
FT                   /product= "Human pyruvate dehydrogenase complex E2"
FT     variation     replace (857, C)
FT                   /*tag= c
FT                   /note= "This variation occurs during polymorphism"
FT     variation     replace (1255, A)
FT                   /*tag= d
FT                   /note= "This variation occurs during polymorphism"
XX
PN     EP1118679-Al.
XX
PD     25-JUL-2001.
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
XX
PF   12-JAN-2001; 2001EP-0300271.
PR   18-JAN-2000; 2000GB-0000992.
XX
PA   (ASTR ) ASTRAZENECA AB.
XX
PI   Smith JC, Anand R, Morten JEN;
XX
DR   WPI; 2001-434704/47.
DR   P-PSDB; AAE05087.
XX
PT   Diagnosis of a polymorphism in the pyruvate dehydrogenase complex E2
PT   useful for, e.g., linkage study and haplotyping comprises determining
PT   the sequence at one or more of positions 857 and 1255 in the PDH E2
PT   gene.
XX
PS   Claim 3; Page 12-13; 18pp; English.
XX
CC   The invention relates to polymorphisms in human pyruvate dehydrogenase
CC   complex E2 (PDH E2 or PDC E2) and its corresponding DNA molecule. PDH E2
CC   polymorphism is used in the diagnosis and treatment of diseases in which
CC   modulation of pyruvate dehydrogenase activity is of therapeutic benefit,
CC   such as diabetes, obesity, asthma, sepsis, peripheral vascular disease,
CC   muscle ataxia, weakness, myocardial ischaemia, lactic acidaemia, coronary
CC   failure and certain cardiac myopathies. The invention also relates to
CC   method for diagnosing a polymorphism in human PDH E2 gene. PDH drugs are
CC   used in the preparation of medicament for treating PDH-mediated diseases
CC   in humans. Polymorphisms are used as a genetic marker in a linkage study
CC   and also in bioinformatic analysis selected from homology searching,
CC   mapping, haplotyping, genotyping and pharmacogenetic analysis.
CC   The present DNA sequence encodes human pyruvate dehydrogenase complex E2
CC   (PDH E2 or PDC E2) protein.
XX
SQ   Sequence 2583 BP; 742 A; 562 C; 564 G; 715 T; 0 other;
```

```
Query Match             100.0%;  Score 1001;  DB 22;    Length 2583;
Best Local Similarity   100.0%;
Matches 1001; Conservative   0; Mismatches   0; Indels   0; Gaps   0;
```

```
Qy          1 atatatcaagcatttgctgactataggccaaccgaagtaacagattaaaaacacacagtgc 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 atatatcaagcatttgctgactataggccaaccgaagtaacagattaaaaacacacagtgc 859

Qy         61 caccacctacccaccccaccccggtggccgctgttcctccaactcccagccttagctccta 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 caccacctacccaccccaccccggtggccgctgttcctccaactcccagccttagctccta 919

Qy        121 cacctcaagtcacccctgcccagctactcctgctggaccaaaagggaaagggtgtgttagcc 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 cacctcaagtcacccctgcccagctactcctgctggaccaaaagggaaagggtgtgttagcc 979

Qy        181 ctcttgcaaagagagttggcaatagagaaggattgatcttacacacaagtaaaagggacag 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 ctcttgcaaagagagttggcaatagagaaggattgatcttacacacaagtaaaagggacag 1039

Qy        241 gaccagaatggtagaatcaccaagaaagaatcagactcttttgtgcctagtaaagttgctc 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 gaccagaatggtagaatcaccaagaaagaatcagactcttttgtgcctagtaaagttgctc 1099

Qy        301 ctgctccggccagctgttgtgtgtgccaccaagttcctctacaggtg 360
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ctgctccggccagctgttgtgtgtgccaccaagttcctctacaggtg 1159

Qy        361 tcttcacagatatcccaatcagcaacattcgtcggttattgcacagcgattaatgcaat 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1160 tcttcacagatatcccaatcagcaacattcgtcggttattgcacagcgattaatgcaat 1219

Qy        421 caaagcaaaccatacctcattattaccttctatcgatgtaaatatgggagagaagtttgt 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1220 caaagcaaaccatacctcattattaccttctatcgatgtaaatatgggagagaagtttgt 1279
```

```
Qy        481 tggtacggaaagaaacttaataatagatattagagagggggaagcaaaaattctgtcaatgact 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1280 tggtacggaaagaaacttaataatagatattagagagggggaagcaaaaattctgtcaatgact 1339

Qy        541 tc.tcataaaagctcagcttggcatgtttaaaagtccgaagcaaaattcttcttgga 600
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1340 tcatcataaaagctcagcttggcatgtttaaaagtccgaagcaaaattcttcttgga 1399

Qy        601 tggacacagttataagacaaatcatgttgttgtagtcagtgttcggttcgtagtactctg 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1400 tggacacagttataagacaaatcatgttgttgtagtcagtgttcggttcgtagtactctg 1459

Qy        661 caggactcatcacacctattgtgttaatgcacatataaaagagagttggaaacacatgcta 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1460 caggactcatcacacctattgtgttaatgcacatataaaagagagttggaaacacatgcta 1519

Qy        721 atgatgttgttcttagcaaccaaagcaagagaggtaaactacagccacatgaattcc 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1520 atgatgttgttcttagcaaccaaagcaagagaggtaaactacagccacatgaattcc 1579

Qy        781 agggtggcacttttacatctccaattagtgaatgtttggaattatctctgcta 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1580 agggtggcacttttacatctccaattagtgaatgtttggaattatctctgcta 1639

Qy        841 tattaaccacctcaagcatgtgtatttggcaattgtgtcgcttcagaggatataaactgtcc 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1640 tattaaccacctcaagcatgtgtatttggcaattgtgtcgcttcagaggatataaactgtcc 1699

Qy        901 ctgcagataatgaaaaaggggtttgatgtggctagcatgatgtctgttacactcagttgtg 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1700 ctgcagataatgaaaaaggggtttgatgtggctagcatgatgtctgttacactcagttgtg 1759

Qy        961 atcaccgggtggtgatggagcagttggagcccagtggctt 1001
              |||||||||||||||||||||||||||||||||||||||||
Db       1760 atcaccgggtggtgatggagcagttggagcccagtggctt 1800
```

```
RESULT    2
AAF21916
ID   AAF21916 standard; DNA; 3859 BP.
XX
AC   AAF21916;
XX
DT   27-MAR-2001 (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 303.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; immunosuppressive;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200005173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;      99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB59013.
XX
```

```
PT   New human breast and ovarian cancer associated gene sequences and the
PT   polypeptides encoded by these genes, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 1; Page 715-716; 1299pp; English.
XX
CC   Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC   proteins AAB58711 - AAB59128. The DNA and Protein sequences are
CC   associated with breast and ovarian cancer. Included in the invention are
CC   sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC   isolation and characterisation of the DNA and protein sequences of the
CC   invention. The breast and ovarian cancer associated DNA, protein, agonist
CC   or antagonist sequences exhibit cytostatic; immunosuppressive;
CC   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC   antibacterial; antifungal; antiparasitic and cardiant activity. The
CC   polynucleotide and protein sequences are used in the diagnosis of cancer,
CC   particularly breast and ovarian cancer. The nucleic acid sequences,
CC   proteins, agonists and agonists may also be used in the diagnosis,
CC   prevention and treatment of immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; cardiovascular disorders such as
CC   myocardial ischaemias; wound healing; neurological diseases such as
CC   cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ   Sequence 3859 BP; 1162 A; 760 C; 820 G; 1106 T; 11 other;

Query Match             99.9%;  Score 1000.2;  DB 21;  Length 3859;
Best Local Similarity   99.8%;  Pred. No. 3.3e-278;
Matches   999;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 atatatacagatcattttgctgactataggcaaccgaagtaacagagatttaaaaccaagtgc 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| ||
Db       1069 atatatacagatcattttgctgactataggccaaccgaagtaacagagatttaaaaaccaaggyc 1128

Qy         61 caccacactacccacccccccgtggccgtgttcctccaacctccccgccttagctccta 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1129 caccacctaccccaaccccctccaactcccccgccttagctccta 1188

Qy        121 caccttcagcacctgcccagctactcctgctctggaccaaaggagaaggggtgtgttgttagcc 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1189 caccttcagcacctgcccagctactcctgctctggaccaaaggagaaggggtgtgttgttagcc 1248

Qy        181 ctcttgcaaagaagttggcagtagagaaagggattgatcttcacacaagtaaaaggacag 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1249 ctcttgcaaagaagttggcagtagagaaagggattgatcttcacacaagtaaaaggacag 1308

Qy        241 gaccagatggtagaatacaccaagaaggatatcgaactctttgtgcctagtaagttgctc 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1309 gaccagatggtagaatacaccaagaaggatatcgaactctttgtgcctagtaagttgctc 1368

Qy        301 ctgctccggccagctgtgtgtgtgcctccaagggtcctggaatggcaccagttcctcacaggtg 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 ctgctccggccagctgtgtgtgtgcctccaagggtcctggaatggcaccagttcctcacaggtg 1428

Qy        361 tcttcacagatatcccaatcagcaacattcgtcggttattgcacacgcgatcaatgcaat 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 tcttcacagatatcccaatcagcaacattcgtcggttattgcacacgcgatcaatgcaat 1488

Qy        421 caaagcaaaccatacctacctcattattacttcatctcgatgtaaatatgggaagaagtttgt 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 caaagcaaaccatacctacctcattattacttcatctcgatgtaaatatgggaagaagtttgt 1548

Qy        481 tggtacggaaagaacttaataagatattagaaggagaagcaaaaattcctgtcaatgact 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 tggtacggaaagaacttaataagatattagaaggagaagcaaaaattcctgtcaatgact 1608

Qy        541 tcatcataaaagcttcagcttgggcatgttaaaagtcccgaagcaaatctcctggga 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 tcatcataaaagcttcagcttgggcatgttaaaagtcccgaagcaaatctcctgga 1668

Qy        601 tggacacagttatataagacaaaatcatgttgttgatgtgtcagttgtgcggtcagtactcctg 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 tggacacagttatataagacaaaatcatgttgttgatgtgtcagttgtgcggtcagtactcctg 1728

Qy        661 caggactcatcacacctattgtgttcacagttataatgcacatataaaaggagtggaaaccattgcta 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 caggactcatcacacctattgtgttcacagttataatgcacatataaaaggagtggaaaccattgcta 1788

Qy        721 ataatgtgttcttagctacaaacagaaggcaagagagaggggtaaactacacagccacatgaattcc 780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 atgatgttgttcttagctacaaacagaaggcaagagagaggggtaaactacacagccacatgaattcc 1848

Qy        781 agggtggccacttttacgatctccaattagttggaattagttaagaattctctgcta 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 agggtggccacttttacgatctccaattagttggaattagttaagaattctctgcta 1908

Qy        841 ttattaaccccaccaagcatgtattttggcaattggcttcaagtaggatgagaggatttgtcc 900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 ttattaaccccaccaagcatgtattttggcaattggcttcaagtaggatgagaggatttgtcc 1968

Qy        901 ctgcagataatgaaaaagggtttgatgtggctagacatgatgtctgttctacactcagttgtg 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 ctgcagataatgaaaaagggtttgatgtggctagacatgatgtctgttctacactcagttgtg 2028

Qy        961 atcaccgggtgggtggatggagcagttggagcccagtggctt 1001
              ||||||||||||||||||||||||||||||||||||||||||
Db       2029 atcaccgggtgggtggatggagcagttggagcccagtggctt 2069

RESULT 3
AANB2410
ID   AANB2410 standard; DNA; 1370 BP.
XX
AC   AANB2410;
XX
DT   23-NOV-1990 (first entry)
XX
DE   Plasmid pRMIT encoding 70 kD mitochondrial autoantigen of primary
DE   biliary cirrhosis.
XX
KW
OS   Rattus rattus.
XX
PN   WO8804689-A.
XX
PD   30-JUN-1988.
XX
PF   16-DEC-1987;    87WO-AU00427.
XX
PR   16-DEC-1986;    86AU-0009523.
XX
PA   (AMRA-) AMRAD CORP LTD.
PA   (COPP/) COPPEL RL.
XX
PI   Coppel RL, Gershwin ME;
XX
DR   WPI; 1988-190632/27.
DR   P-PSDB; AAP82931.
XX
PS   Claim 1; Fig 6; 42pp; English.
XX
PT   Novel DNA encoding primary biliary cirrhosis autoantigen - used to
PT   express polypeptide useful in diagnostic tests and for treatment of
PT   prim biliary cirrhosis.
XX
CC   The clone was isolated from a rat liver cDNA lambda gt 11-Amp3
CC   library. It expresses an antigen which is specifically reactive
CC   with sera from patients with primary biliary cirrhosis (PBC). The
CC   clone does not contain the full-length sequence of the antigen. The
CC   sequence is not present in mitochondrial DNA; the 70 kD protein is
CC   therefore coded for by nuclear genes. The protein can be used in
```

```
CC   highly sensitive ELISA or to treat PBC.
CC   See also AAN82411.
XX
SQ   Sequence 1370 BP; 368 A; 349 C; 334 G; 319 T; 0 other;

Query Match           73.5%;  Score 735.8; DB 9;  Length 1370;
Best Local Similarity 85.6%;  Pred. No. 4.5e-202;
Matches 859;  Conservative 0; Mismatches 127; Indels 17; Gaps 3;

Qy    1 atatatcagcatttgctgactataggccaacgaagtaacagatttaaaaccaacaagtgc 60
        |||| |||||||||| | |||| | || ||||||||||| |||| ||||||| ||||| |
Db  380 atatagcagcatttgcagactacagcggccaacgaagtgacaagcttaaagccacagcac 439

Qy   61 caccacctaccccacccccggtggccgtgttcctccaactcccagctttagctccta 120
        |||||||| | ||||||||  | |||| |||| |||||| ||||| |||||| ||| ||
Db  440 caccacctgtcccaccccggcagctgtcctcctccaatcccccagccttagcacctta 499

Qy  121 caccttcagcaccctgccagctacctctgtggacaaagaaagggaaggtgttgttagcc 180
        |  || |      |||    || |  |  | |||||||||||||||||||| ||| ||||
Db  500 cccc----------ctcagcgccctgcctctgctggacaaagaaagggaaggtgttcgttagcc 547

Qy  181 ctcttgcaaagaagttggcagtagagaaagggattgaagaagtaaaagggacag 240
        |||||||||||||| |||||| |||||||||||||| |  || |||||||||||
Db  548 ctcttgcaaagaaattgcagcagagaaagggattgaactgaagctaaagggacg 607

Qy  241 gaccagatggtagaatcaccaagaaggaagatatcgacctctttgtcctcagtaaagttgctc 300
        ||||||| | |||||||| | |||| ||||  |  ||||||||| ||||| ||   |||||||
Db  608 gaccagaggcaggaaatcatcaagaggacatggacctctttttgtcctaggctgctc 667

Qy  301 ctgctccggcagctgttgtcctcccacagtcctggaatggcaccagttcctacaagtg 360
        ||||| ||| || |||| ||||||||| ||| ||||  |||||||| |||| ||||||
Db  668 ctgcggctgcagctg---ctcccccgggtccaagagtggcaccaactcctgcagtg 724

Qy  361 tcttcacagatatcccatcagcacacattctcggttattcgcacagcgattaatgcaat 420
        |||||  ||| ||||| |||  |  |||| |  |  |||| ||||  |||| |  |||||
Db  725 tcttcatcagacatcccccaccccattccattcgtcgatggtaaatatgggagaagtgcagt 784

Qy  421 caaagcaaacatacctcattattacctttcatgatgtaaaatatgggagaagtttttgt 480
        | ||||| ||| |||||||||||||||||  | |||||||||| ||||| |||  |||
Db  785 cgaagcagactatacctcattattacctttctgtgatgtaaatatggagagtgctgt 844

Qy  481 tgtacggaaagaacttaataagatattagaagggagaagcaaaaatttcctgtcaatgact 540
        ||  | || ||||||||||||| | | || ||| |||||  ||||||| ||||||| ||||
Db  845 tggttgcgaaggaacttaataagtcttgaaggtaaagaaaatcccgtcaaatgact 904

Qy  541 tcatcataaagcttcagcttggcatgttaaagttcccgaagcaagcaaattctcttcttgga 600
        ||||||||||||||||||||||| ||  |||| ||||  |||||||| || |||||| ||||
Db  905 tcatcataaagcttcagcttggcctgaagtaactcatcttgga 964

Qy  601 tggaacagagttatagagacaaatcatgttgttgtgcggtcagtgttactcctg 660
        |||| |||||| |||||||||||||||  || |||| ||| ||||| || |||
Db  965 tggacaaagtgtatacgacaaatcatcggttgattgaatgtcagtgttcagtgttactcccg 1024

Qy  661 caggacttcatcaactcattgtgttaatgcacatatataaaagagtggaaaccattgcta 720
        |||||||| ||||| | ||||| ||||||||||| ||| ||||||| |||||||||||||
Db 1025 caggacttatcatcaccctattgttaatgcacacataaaagagactgggaaaccattgcta 1084

Qy  721 atgaatgttgtttcttagcacaacgaagtacagcgcacatgaatcc 780
        | ||||||||||||||||||   |  |||||| ||| ||  || ||
Db 1085 gtgatgttgtttcttagccctccaaagcaagctaaacttcagctcacgagtcc 1144

Qy  781 agggtggcacttcacgtctctccaatttaggaatgttcgaattaagaattctctgcta 840
        |||||||||| | ||  |||||||||| ||| ||||| ||||||||||||| |||| ||
Db 1145 agggtggcacatttacaatctccaacttagggatgttcggaattaagaatttctctcgga 1204

Qy  841 ttattaacccaccacctcaagacgtgtattttggcaattggtgcttcagaggataaactggtcc 900
        |||||||||||||||||||||| |||||||||||||||||| ||| | ||||||||||| ||||
Db 1205 ttattaacccaccacctcaaggatgtattttggcaattggtcttccgaggataaactgatcc 1264

Qy  901 ctgcagataatgaaaaagggtttgatgtgggtagcataatgtcgt--tacactcagttg 958
        ||||||||||||| | |||||| || ||| | ||  |  |  |||  |||| |||| ||
Db 1265 ctgcagataatgagagaaaaagggttgacgtggctagtgtctcacacaccaagctg 1324

Qy  959 tgatcaccggttggtgggatggagcagttgggagccagttggaagcccagtggctt 1001
        ||||| | ||| ||| |||||||| |||| ||||| || ||| ||| ||||||||||
Db 1325 tgatcatcgagttgtggatggagccagttggagccagttggagccccagtggctt 1367
```

```
Qy  959 tgatcaccggttggtgggatggagcagttgggagccagttggaagcccagtggctt 1001
        ||||| | ||| ||| |||||||| |||| ||||| || ||| ||| ||||||||||
Db 1325 tgatcatcgagttgtggatggagccagttggagccagttggagccccagtggctt 1367

RESULT  4
AAN82411
ID   AAN82411 standard; DNA; 966 BP.
XX
AC   AAN82411;
XX
DT   23-NOV-1990  (first entry)
XX
DE   Insert from cDNA plasmid encoding autoantigen of primary biliary
DE   cirrhosis.
XX
KW   Primary biliary cirrhosis autoantigen; ss.
XX
OS   Homo sapiens.
XX
PN   WO8804689-A.
XX
PD   30-JUN-1988.
XX
PF   16-DEC-1987; 87WO-AU00427.
XX
PR   16-DEC-1986; 86AU-0009523.
XX
PA   (AMRA-) AMRAD CORP LTD.
PA   (COPP/) COPPEL RL.
XX
PI   Coppel RL, Gershwin ME;
XX
DR   WPI; 1988-190632/27.
DR   P-PSDB; AAP82932.
XX
PT   Novel DNA encoding primary biliary cirrhosis autoantigen - used to
PT   express polypeptide useful in diagnostic tests and for treatment of
PT   prim biliary cirrhosis.
XX
PS   Claim 1; Fig 8; 42pp; English.
XX
CC   The clone was isolated from a human placental cDNA lambda gt11-Amp3
CC   library. It was obtd. by screening using pMIT (AAN82410) as a probe.
CC   The clone expresses an antigen which is specifically reactive with
CC   sera from patients with primary biliary cirrhosis (PBC). The
CC   protein can be used in highly sensitive ELISA to detect anti-mito-
CC   chondrial antibodies or autoreactive cells.
CC   See also AAN82410.
XX
SQ   Sequence 966 BP; 289 A; 230 C; 207 G; 240 T; 0 other;

Query Match           69.9%;  Score 699.4; DB 9;  Length 966;
Best Local Similarity 99.9%;  Pred. No. 1.2e-191;
Matches 700;  Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy    1 atatatcagcatttgctgactataggccaacgaagtaacagatttaaaaccaacaagtgc 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  266 atatatcagcatttgctgactataggccaacgaagtaacagatttaaaaccaacaagtgc 325

Qy   61 caccacctaccccacccccggtggccgtgttcctccaactcccagctttagctccta 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  326 caccacctaccccaccccccggtggccgtgttcctccaactcccagctttagctccta 385

Qy  121 caccttcagcaccctgccagctacctctgtggacaaagaaagggaaggtgttgttagcc 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  386 caccttcagcaccctgccagctacctctgtggacaaagaaagggaaggtgttgttagcc 445

Qy  181 ctcttgcaaagaagttggcagtagagaaagggattgatctataccaagttgatttttagc 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  446 ctcttgcaaagaagttggcagtagagaaagggattgatctataccaagttgatttttagc 505
```

Qy   241 gaccagatggtagaatccacccagaagaaggatatcgactccttttgtgcctagtaaagttgctc 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   506 gaccagatggtagaatccacccagaagaaggatatcgactccttttgtgcctagtaaagttgctc 565

Qy   301 ctgctccggcagctgttgtgctcccacagttcctacagtg 360
          ||||||||||||||||||||||||||||||||||||||||
Db   566 ctgctccggcagctgttgtgctcccacaggtccacagttcctacagtg 625

Qy   361 tcttcacagaatccagcaaattcctgcggttatgcacgcgttatatgcaat 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   626 tcttcacagatatcccaatccagcaacattcgcgggttattgcacagcgattaatgcaat 685

Qy   421 caaagcaaaccataccctcattattacctttctatcgatgtaaatatgggagaagtttgt 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   686 caaagcaaaccataccctcattattacctttctatcgatgtaaatatgggagaagtttgt 745

Qy   481 tggtacggaaagaaacttaataagatattagaaggagaagcaaaaattctgtcaatgact 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   746 tggtacggaaagaaacttaataagatattagaaggagaagcaaaaattctgtcaattgct 805

Qy   541 tcatcataaaagcttcagcttggcatgttaaaagttcccgaagcaaaatctcttctgga 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   806 tcatcataaaagcttcagcttggcatgttaaaagttcccgaagcaaaatctcttctgga 865

Qy   601 tggacacagttataagacaaaatcatgttgttgatgtcagtgtgtcagtacgtcagttcctg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   866 tggacacagttataagacaaaatcatgttgttgatgtcagtgtgtcggtcagttcgg tactcctg 925

Qy   661 caggactcatcacacctattgtgttaatgcacatataaaa 701
          ||||||||||||||||||||||||||||||||||||||||
Db   926 caggactcatcacacctattgtgttaatgcacatataaaa 966

RESULT 5
AAS70149
ID   AAS70149 standard; cDNA; 360 BP.
XX
AC   AAS70149;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #5953.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 5953; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and

CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (I) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II) and are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 360 BP; 106 A; 97 C; 81 G; 76 T; 0 other;

Query Match   25.0%; Score 250.4; DB 23; Length 360;
Best Local Similarity   99.6%; Pred. NO. 3.6e-62;
Matches 251; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 atatatcagattgctgactataggccaaccgaagtaacaagatttaaaacaacaagtgc 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   109 atatatcagattgctgactataggccaaccgaagtaacaagatttaaaacaacaagcgc 168

Qy   61 cacaacctacccaccccccggtgggccgctgttcctcccaactccccagcccttttagcctca 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   169 cacaacctacccaccccccggtgggccgctgttcctcccaactccccagcccttttagctccta 228

Qy   121 cacccttcagcaccctgcccagctactcctgctggaccaaaggagagggtgttgttgttagcc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   229 cacccttcagcaccctgcccagctactcctgctggaccaaaggagagggtgttgttgttagcc 288

Qy   181 ctcttgcaaagaagaagttggcagtaagagaagggattgatcttacacaagtaaaaacacaagtcag 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   289 ctcttgcaaagaagaagttggcagtaagagaagggattgatcttacacaagtaaaaacacaagcgc 348

Qy   241 gaccagatggta 252
          ||||||||||||
Db   349 gaccagatggta 360

RESULT 6
ABI99637
ID   ABI99637 standard; cDNA; 319 BP.
XX
AC   ABI99637;
XX
DT   07-MAR-2002 (first entry)
XX
DE   Mouse ischaemic condition related cDNA sequence SEQ ID NO:662.
XX
KW   Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW   vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS   Mus musculus.
XX
PN   WO200188188-A2.
XX
PD   22-NOV-2001.
XX
DT   18-MAY-2001; 2001WO-JP04192.
XX
PR   18-MAY-2000; 2000JP-0145977.
XX
PA   (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

```
XX
PI   Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
DR   WPI; 2002-034733/04.
XX
PT   Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT   expression levels of particular genes defined in the specification or
PT   by determining the expression profile of a gene group comprising these
PT   genes -
XX
PS   Claim 2; Page 1660; 2690pp; English.
XX

CC   The present invention describes a method for examining ischemic
CC   conditions, comprising measuring the expression levels of particular
CC   genes (I) in a test sample or determining the expression profile of a
CC   gene group in the sample comprising genes selected from (I). The method
CC   is useful for examining the ischaemic condition (e.g. comprehensive
CC   ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC   expression levels of particular genes (ABI99202 to ABI99912, encoding
CC   the protein sequences in ABB57020 to ABB57374) or by determining the
CC   expression profile of a gene group comprising these genes. The
CC   expression levels or expression profiles produced by these genes are
CC   used as an indicator when screening for ischaemic condition-improving
CC   drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC   represent PCR primers for a mouse ischemic condition related sequence,
CC   which are used in the exemplification of the present invention.
XX
SQ   Sequence 319 BP; 89 A; 60 C; 77 G; 93 T; 0 other;

Query Match             22.3%;  Score 223;  DB 24;  Length 319;
Best Local Similarity   88.9%;  Pred. No. 2.8e-54;
Matches  241;  Conservative  0;  Mismatches  30;  Indels  0;  Gaps  0;

Qy  731 ttcttagcaaccaaagcaaggcaggagggggtaaactacagccacatgaattccaggtggcac 790
         ||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    1 ttctttagccctccaaagcgagagaggggtaaacttagccctcaatgagttccaggtggaac 60

Qy  791 tttacgatctccaattaggaatgttggaattaagaattctctgcttattattaaacc 850
         ||||| |||| || ||||||| |||||||||||| || |||||||||| ||
Db   61 attacaaatctccaacttaggatgtgttggaattaagaattctctgcaattagtagtcc 120

Qy  851 aacctcaagcatgtattttggcaaatggttgctccagaggaataaactggtccctgcagataa 910
         ||||| |||||| |||| ||||||||||||||||||| ||||||||| ||||| ||||||||
Db  121 aacctcaggcatgtattctggcaatccggtgctctcataggataaactgaatccagcagataa 180

Qy  911 tgaaaaaggtttgatgtgtggctagcatgcaatgatgatgtctcttttacactcagttgttga 970
         || ||||||||| || |||||||||||  |||||||||  ||||||| ||||||||||||
Db  181 tgagaaaggcttgatggtggctagtgcctgtccgttacactcatcatttgtctgatcatcacc 240

Qy  971 ggtggatggagcagttggagcccagtggctt 1001
         ||||||||||||||||||||||||||||||
Db  241 tgtggatggagcagttggagcccagtggctt 271
```

```
RESULT 7

ID   ABL20587 standard; DNA; 1502 BP.
XX
AC   ABL20587;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 13234.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
OS   Drosophila melanogaster.
XX
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
```

```
XX
PF
XX
PR   23-MAR-2001; 2001WO-US09231.
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PS   Claim 1; SEQ ID NO 13234; 21pp + Sequence Listing; English.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX

CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511), and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ   Sequence 1502 BP; 355 A; 435 C; 413 G; 299 T; 0 other;

Query Match             21.0%;  Score 209.8;  DB 23;  Length 1502;
Best Local Similarity   54.3%;  Pred. No. 4.2e-50;
Matches  527;  Conservative  0;  Mismatches  417;  Indels  27;  Gaps  4;

Qy   36 gtaacagatttaaaacaccacacagtgcaccacctacccacacccccggtggccgcgtgttcct 95
         | |||| | |||   || ||||| |  |  || | ||| ||||| || |  |||  |||
Db  504 gcagcagctccagctccagcaccgcagcagcagctccgccaccacacaccc 563

Qy   96 ccaactccccagcctttagctcctacacacttcagcacctgcccagctactcctgctgga 155
         || ||| | || || ||| | || | | | ||| | || || |||| ||  |||||
Db  564 gccgcagctccagcagccgctgcgctcccaccagctccgcctgcttggcagttgag 623

Qy  156 ccaaagggaagggtgttgttagccctccttgcaaagaagttggcagtagagaaaggatt 215
         || |||| |  || | | || || |  ||| || ||| | | || | | ||| |||
Db  624 ggtacgggccgtgtatatgccagtccaatggccaagtgctcagcagccgaggcacaacagctg 683

Qy  216 gatcttacacagtaaacagggagcagagcccaagtgggtagaatacacaagagaatacgac 275
         ||| ||| | |  || ||| | || | |||| ||||| || |  ||| || | |||
Db  684 cgtct---acaagcaaggcagtgaagtccaagctccaattaaatctggatcttgca 740

Qy  276 tctttgtgcctagtaaagttgctcctgctccggcacgctggttgtgcctcccacagtcct 335
         ||| |||  ||| | |||| ||||| || ||| | |||| |  || | ||| |||| ||
Db  741 ggccagaaagcagcaagccagctggtcggggcacctgcccaaggcgccccagtggcagct 800

Qy  336 ggaatggcccagtcctcacaggtgtcttcacagatatcccaatcagcaacattctcg 395
         ||| || ||||| |  |||| || | |  ||||| | |||| || | ||| || |
Db  801 ggagcgcg-------------ctacagaggacattccggtgacaaactgacttgccgca 842

Qy  396 gtaattgcacacggatataatgcaatccaaagcaaaccatcaccattattacctttctatc 455
         || ||| ||| ||| | |  | | || ||||| | || | || |||| ||| ||||
Db  843 gtgattgccaaaacgtctgctgaggtccaagacaacacctcactactacgtcaccgtg 902

Qy  456 gatcgtaaatatgggagaagtttcgttggtacggaagaacttaataagatattagaaggg 515
         || || || | ||| | | | || | ||| | |  ||  | ||| || | | |||  |
Db  903 caatgccaagtgataagctgctgaagttccgccgcgaagttgaacaagaagtacgagaag 962

Qy  516 agaag---caaaatttctgtcaatgacttcatcataaaagcttcagcttttgcatgtta 572
         ||||    ||||| | || ||| | || |  ||| |  || ||| || |   | | |
Db  963 caggtgcccgagcgtctctgtaaacgacttcattatcaaggcgcgtggccattgccagtctt 1022
```

```
Qy     573 aaagttcccgaagcaaatcttcttggatggatggaacaagtgtataagacaaaatcatgttgtt 632
              ||| ||||||| || ||  ||||| ||||| ||||| ||| |||   ||||| |||| ||||
Db    1023 aagttcccggaagccgaactccgcctggatggacaagttactgcaagtacgacgcgtc 1082

Qy     633 gatgcaagttgttcggtcagtcactcctcggagactcatcacacctattgtgttaatgca 692
              |||| ||||| ||||||  ||| || ||||||||| ||||||||||||| |||||||||
Db    1083 gatgttcggttgctgttctccacacaagggtctgattacccgattgtttcaatgcc 1142

Qy     693 catataaaaggaagtggaaacaattgtataatgatgttgtttttcttagcaacacaaagcaaga 752
              ||  || |||||| ||| || |||||  | ||||||||||||||||||||||||||||||||
Db    1143 gaccggcaagggtgtcctggagatctctaaggatgtcaaggcgctggcagccaagcggcgc 1202

Qy     753 gagggtaaactacacaccgaattctccaatctccaatttagga 812
              |||||| |||||||   |||||| ||| ||| | |||||||||
Db    1203 gacaacaaacttcagcccacgaattccaggtggcaccactctcggtttctaatctggc 1262

Qy     813 atgtttggaattaagaaaattctctgatattattaaccactcaagcacatgtatttggca 872
              |||| ||||||||||||| ||||||| ||||||| ||||||||||||||| ||||| |||
Db    1263 atgtcggttgtgaaccagttcgcccctgtcatcaacacccgtgattacccgtgtttcaatgcc 1322

Qy     873 attggtgcttcagaggagataaactggttccctg---cagataatgaaaaagttgatgtg 929
              ||||| ||||||||||| || |||||||  |||   |||||||||||  | ||||| ||
Db    1323 attggcaccacacgaaacaagttggttgctgatcccggactctcaagggttcaaggag 1382

Qy     930 gctagcatgatgtctgttacactcaagttgtgatcaccgggtggtggagcagttgga 989
              ||||| ||||| || || || | |||| |||||| |||| |||||||||||||||
Db    1383 gtcaacatgctgacgtgacgtcaccctgagtcgatcatccgtgttggatggtgctgttgct 1442

Qy     990 gcccagtggct 1000
              ||| ||||||
Db    1443 gccaggtggct 1453

RESULT 8
ABL09105
ID  ABL09105 standard; cDNA; 1850 BP.
XX
AC  ABL09105;
XX
DT  26-MAR-2002  (first entry)
XX
DE  Drosophila melanogaster expressed polynucleotide SEQ ID NO 21797.
XX
KW  Drosophila; developmental biology; cell signalling; insecticide;
KW  pharmaceutical; gene; ss.
XX
OS  Drosophila melanogaster.
XX
XX
CC  WO20017142-A2.
PD  27-SEP-2001.
XX
PF  23-MAR-2001; 2001WO-US09231.
XX
PR  23-MAR-2000; 2000US-191637P.
PR  11-JUL-2000; 2000US-0614150.
XX
PA  (PEKE ) PE CORP NY.
XX
PI  Venter JC, Adams M, Li PWD, Myers EW;
XX
DR  WPI; 2001-656860/75.
DR  P-PSDB; ABB65002.
XX
PT  New isolated nucleic acid detection reagent for detecting 1000 or more
PT  genes from Drosophila and for elucidating cell signalling and cell-cell
PT  interactions -
XX
PS  Claim 1; SEQ ID NO 21797; 21pp + Sequence Listing; English.
XX
CC  The invention relates to an isolated nucleic acid detection reagent
CC  capable of detecting 1000 or more genes from Drosophila. The invention is
CC  useful in developmental biology and in elucidating cell signalling and
```

```
CC  cell-cell interactions in higher eukaryotes for the development of
CC  insecticides, therapeutics and pharmaceutical drugs. The invention
CC  discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC  sequences (ABL01040-ABL16175) and the encoded proteins
CC  (ABB57/37-ABB72072).
CC  The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 1850 BP; 430 A; 542 C; 485 G; 393 T; 0 other;

Query Match        21.0%;  Score 209.8;  DB 23;  Length 1850;
Best Local Similarity    54.3%;
Matches 527;  Conservative  0;  Mismatches 417;  Indels 27;  Gaps  4;

Qy      36 gtaacagatttaaaccacacaagtgccacacttagcacaccacccccccggtggccgctgtcct 95
              || || |||| |||  |||| |||||| ||  ||||| ||||| |||||||||  |||||| ||
Db     852 gcagcagctccagctccagcacccggcacgcagcagcagcagcagtaccaccacaccacacccc 911

Qy      96 ccaactccccagcccttagctcctacacacttcagcacacttcagcacactcctgctctgga 155
              ||||| ||||||| ||||||||| |||   ||||||||| ||||||||| ||||||| ||||
Db     912 gccgcagctccagcagccgcctccaccaccagctccccagcctccaaagcagaaaaagggatt 971

Qy     156 ccaaaggggaagggttgttgttagccccttgcaaaagaagttgcagtagagaaaagggatt 215
              ||||| |||||||||||||| || |||   |||||||||||||| ||| ||||| ||||||
Db     972 ggtacggccgtgtatatgcaggtccagtgccaagtgccagctagagcgagggcacaacagctg 1031

Qy     216 gatcttacacaagtaaaaggaacaagatggtagaatcaccagaagaagatatcgac 275
              ||||||| || || |||| |||  |||| ||||| |||||||||||| || ||
Db    1032 cgtct--acaagcaagggacgtggagtccaattaaatctggtgatcttgca 1088

Qy     276 tcttttgtgccttagtaaagttgctcccctcgctcccgcagcgtgtgttgctccccacaggtcct 335
              ||||||||| || |||||| ||||| ||||||||||| |||||||||||||||| |||||| ||
Db    1089 ggcagaacaagcagcagcagcccagctgccgtgcggccacccgccaagcccccaggcagct 1148

Qy     336 ggaatggcaccagttcctacaggtgtcttcacagatatcccaatcagcaacattcgtcgg 395
              |||| ||||||||||| | ||||||||||| ||| ||| ||||| || |||||| |||
Db    1149 gggcgcg--------ctacggagacattccggtgacacaacaacttcgcgcg 1190

Qy     396 gttattgcacacgcgattaatgcaatcaaagcaaaccacatacccatcattattacctcatc 455
              |||||||||  || || |||| ||| || ||||| |||| |||||||||||||| ||| |||
Db    1191 gtgattgcaaacgtctgctggagtccaagacacacaactgcctcactactacgtcacgtg 1250

Qy     456 gatgtaaataggagaagagttgttgtggtacgagaaagaaacttaataagatattagaagg 515
              |||| |||| |||| ||| |||||| |||||||| |||||||  ||||| | |||||||| |
Db    1251 caatgccaagtggataagctgctgaagttccgcgcaggagtgaacagagaagtacgagaag 1310

Qy     516 agaag---caaaatttctgtcaatgacttcaatcataaagctcaagcttgcatgtta 572
              | |||   |||||||||| | |||| ||||||| || |||||||||| ||| ||| ||
Db    1311 cagggtgcccgcgtctctgtaaacgactcaatatctcaaggccggtggccattgcaagtctt 1370

Qy     573 aaagttcccgaagcaaatcttcttggatggatggaacaagtgtataagacaaaatcatgtgtt 632
              |||||||| ||||| |||||| || ||||| |||||||||| ||||   ||||| |||| |||
Db    1371 aagttcccggaagccgaactccgcctggatggacacagtaattgcaagtacgacgcgtc 1430

Qy     633 gatgtcagttgttcggtcagtcactcctcgcagactcatcacacctattgtgttaatgca 692
              |||| |||||| ||||||  ||| || ||  |||| ||||||||||||| |||||||||
Db    1431 gatgttcggttgctgttctccacagcaaggggtctgattacccgattgtttcaatgcc 1490

Qy     693 catataaaaggaagtggaaaccattgctaatgatgttgtttttcttagcaacacaaagcaaga 752
              ||  || |||||| ||| || |||||| | ||||||||||||||||||||||||||||||||
Db    1491 gaccggcaagggtgtcctggagatctctaaggatgtcaaggcgctggcagccaagcggcgc 1550

Qy     753 gagggtaaactacacagcccatgaattccaggtggcactttatcgatctccaatttagga 812
              |||||| ||||||| |  |||||||||||||||||||| |||| ||| ||||||||||||
Db    1551 gacaacaaacttcagccccacgaattccaggtggcaccatctcggttctaatctgggc 1610

Qy     813 atgtttggaattaagaaaattctctgtattattaaccactcaagcacatgtatttggca 872
              |||| ||||| ||||||| |||||||| |||||| |||||||||||||| ||||| |||
Db    1611 atgttcggttgtgaaccagtttcgccctgtcatcaacacctcggttctaatctggcc 1670

Qy     873 atgggtgcttcagaggagataaactggtttccctg---cagataatgaaaaggtttgatgtg 929
              ||||| ||||||||||| || ||||||||  |||   |||||||||||  | |||||  ||
Db   ... 
```

```
Db   1671 attggcaccaacgaaacagtggttgttgctgatccgaccagtcctcaaggag 1730
           |||||  |||||  |  |||||  ||||||   |||  |||||
Qy    930 gctagcatgatgtctgttacactcagttgtgatcaccgggttggtggatggagcagttgga 989
           ||||||||||  ||  |  |||  ||||||   ||||||   ||||||
Db   1731 gtcaacaatgctgacgtcacctgagtgctgatcatcgttgttgtggatggtgctgttgct 1790
           |||||  ||   |||||||||   |||||  |||||||  |||
Qy    990 gcccagtgct 1000
           |||  |||||
Db   1791 gccaagtggct 1801
```

RESULT   9
AAS70150
ID   AAS70150 standard; cDNA; 639 BP.
XX
AC   AAS70150;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #5954.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG05963.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
     Claim 1; SEQ ID NO 5954; 103pp; English.

CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 639 BP; 185 A; 130 C; 178 G; 146 T; 0 other;
```
Query Match     15.5%; Score 155.4; DB 23; Length 639;
Best Local Similarity     99.4%;
Matches  156; Conservative   0; Mismatches   1; Indels   0; Gaps   0;
```

```
Qy   79 cggtggccgccgtgttcctccaactccccagccttagctccacacttcagcacctgcc 138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  182 cctgtggccgctgttcctccaactccccagccttagctccacaacttcagcacctgcc 241
Qy  139 cagctactcctgctggaccaaaggaaggggtgtttgttaggccctcttgcaaagaagg 198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  242 cagctactcctgctggaccaaaggaaggggtgtttgttaggccctcttgcaaagaagg 301
Qy  199 cagtagagaaagggattgatcttacacaagtaaaagg 235
          |||||||||||||||||||||||||||||||||||||
Db  302 cagtagagaaagggattgatcttacacaagtaaaagg 338
```

RESULT 10
ABL20586
ID   ABL20586 standard; DNA; 4145 BP.
XX
AC   ABL20586;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 13231.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-1916379P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-658860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 13231; 21pp + Sequence Listing; English.

CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4145 BP; 1165 A; 986 C; 905 G; 1089 T; 0 other;
```
Query Match     12.6%; Score 125.8; DB 23; Length 4145;
Best Local Similarity     63.3%;
Matches  193; Conservative   0; Mismatches  112; Indels   0; Gaps   0;
```

Qy   526 tttctgtcaatgactcaatcatataaaagcttcagcttggcatgttaaaagttcccgaag 585
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2495 tctctgtaaaacgacttcattatcaaggccgtggccattgccagtctaagttcccgaag 2554

Qy   586 caaattcttcttgatggacacagttataagacaaaatcatgttgtgatgtcagtgttg 645
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   2555 cgaactccgcctggatggacacagtaattcgcaagtacgacgtcgatgttcggttg 2614

Qy   646 cggtcagtactcctgcaggactcacacctattgttaatgcacatataaaaggag 705
           |||||||||||||||||||||||||||||||||| ||||
Db   2615 ctgtctccacagacaaggtctgattaacccgattgtttcaatgtcgaccgcaaggtg 2674

Qy   706 tggaaaccattgctaatgatgtgttgttcttagtcacacaaagacaagaagggtaaactac 765
           |||| |||||||||||||||||||||||||||||||||||||||||
Db   2675 tcctggagatctcaaggatgtcaaggcgcgtggacagcaacacaaacttc 2734

Qy   766 agccacatgaattccaggtggtggcacttttacgatctccaattaggaatgtttggaatta 825
           ||||||||||||||||||||||||||||||||||||||||||||
Db   2735 agcccacagaattccaggtggcacaatctcggttctaatctgggcatgttcggtatgg 2794

Qy   826 agaat 830
           |||
Db   2795 gaaat 2799

Query Match   12.6%; Score 125.8; DB 23; Length 4983;
Best Local Similarity   63.3%; Pred. No. 1.5e-25;
Matches 193; Conservative   0; Mismatches 112; Indels   0; Gaps   0;

Qy   526 tttctgtcaatgactcaatcatataaaagcttcagcttggcatgttaaaagttcccgaag 585
           |||||| ||||||||||||||||| ||||||||||||||||||||||||||||||
Db   3333 tctctgtaaaacgacttcattatcaaggccgtggccattgccagtctaagttcccgaag 3392

Qy   586 caaattcttcttgatggacacagttataagacaaaatcatgttgtgatgtcagtgttg 645
           |||||||||||||||||||||||||||||| ||||||||||||||| |||||
Db   3393 cgaactccgcctggatggacacagtaattcgcaagtacgacgtcgatgttcggttg 3452

Qy   646 cggtcagtactcctgcaggactcacacctattgttaatgcacatataaaaggag 705
           ||||||||||||||||||||||||||||| ||||
Db   3453 ctgtctccacagacaaggtctgattaacccgattgtttcaatgtcgaccgcaaggtg 3512

Qy   706 tggaaaccattgctaatgatgtgttgttcttagtcacacaaagacaagaagggtaaactac 765
           |||| |||||||||||||||||||||||||||||||||||||||
Db   3513 tcctggagatctcaaggatgtcaaggcgcgtggacagcaacacaaacttc 3572

Qy   766 agccacatgaattccaggtggtggcacttttacgatctccaattaggaatgtttggaatta 825
           ||||||||||||||||||||||||||||||||||||||||||
Db   3573 agcccacgaattccaggtggcacatctcggttctaatctgggcatgttcggtatgg 3632

Qy   826 agaat 830
           |||
Db   3633 gaaat 3637

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145913.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149722.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0154486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

```
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.

Query Match             53.4%;  Score 125.6;  DB 21;  Length 2348;
Best Local Similarity   12.5%;
Matches  344;  Conservative  0;  Mismatches 279;  Indels 21;  Gaps 3;
                        Pred. No. 1.1e-25;

Qy    368 agatatcccaatcagcacaattcgtcggttgttattgcacagcgaatcaaagca 427
          |||| |||       | | |       |  | | |||||||||| | ||||| ||
Db   1362 agattttccaaatagtcaagatattatagtcaaaacgtttattggaatcaaaaca 1421

Qy    428 aaccatactctcattattcactttctatcgcatgtaaatatgggaagagttgttggtacg 487
          || |  | |  | | ||     ||| |    |||||| ||  ||||||    ||| ||||
Db   1422 gaagataccccattgttattacacatcagatgttgtattggatccctcttctgcatttag 1481

Qy    488 gaaggaacttaataagatattagaaggggaagcaaaatttcctgtcaaatgacttcatcat 547
          |||| ||||| | || |  | |  ||||       | |||||||| ||  ||| ||| | ||
Db   1482 aaaagaacttcaagaaacatggcggtt-----aaagtgtctgtcaatgacatagttat 1535

Qy    548 aaaagttcagcttgcatgttttaaaagttcccgaagcaaattcttcttcttgg-----at 601
          ||||| || || |||  || | |   |   |  |||| || |||| ||| ||     ||
Db   1536 taaagcagtggcagtcgcactgagggaatgtacacaagctaatgcttctctggatgctga 1595

Qy    602 ggacacagttataagacaaaatcatgttgttgatgtcagttcagttcagtactcctgc 661
          || | || || | ||  || |||| ||||| |||| ||| || || ||| || ||||
Db   1596 gaaaggagagatatcgttatgtgtgacagtgttgacatatcaattgcaactgagaa 1655

Qy    662 aggactcatcacacctattgtgttaatgcacatataaaagggagtgaaaccaatgctaa 721
          ||| | ||| |   | ||| |  ||| ||| ||| ||||||   |     |  ||||
Db   1656 gggcttaatgactccaattatcaagaatgcagacagaaatctatttctgttattcctt 1715

Qy    722 tgatgtgttctttagcaaccaagacagagaggtaaactcacagccacatgaattcca 781
          || | | ||| | ||| ||  | || || ||| ||  |  ||| || ||  |||
Db   1716 agaggttaagagagctggacaacaaaagctcgatctggaaagcttgctcctcatgagttca 1775

Qy    782 ggggtggcacctttacgatcccaattaggaatgtttggaattaagaatttctctgctat 841
          |||| |||| |||  | |||  ||| || |||| | ||| || | ||||||||| | ||
Db   1776 aggaggggacaattcagcaatcaaatctagaaatgtatcctgtggtaattctgcgcaat 1835

Qy    842 tattaacccacctcaagacatgtatttggcaattggtgcttcagaggagataaactgtccc 901
          |||||| || ||||| |  | ||  | ||| || ||| |||||||| |  | |||| |||
Db   1836 cattaacccccctcaggctcggggattctacagttcgtagagggaaacaagtggttgagcc 1895

Qy    902 tgcagaataataagaaaagggttga-------tgtggctagcagatgtcgttacac 952
          ||| |  | || | | |||| ||        |||||| ||| | || |  |||||
Db   1896 agtattggactagatggaagcctctgttggacaaaaatgaatgtaacgtt 1955

Qy    953 cagttgtgatcaccggttggttgatggagagcagttggagcccagt 996
          ||||||||||| || |||| |||||||| |||||||||||| | ||
Db   1956 atccgctgatcatcggatcttgatggacaagttggagcttcgt 1999
```

```
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28130.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28139.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 28130; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2595 BP; 967 A; 505 C; 551 G; 572 T; 0 other;
```

```
Query Match             9.7%;  Score 96.6;  DB 23;  Length 2595;
Best Local Similarity   91.9%;
Matches  102;  Conservative  0;  Mismatches  9;  Indels  0;  Gaps  0;
                        Pred. No. 2.8e-17;

Qy    303 gtccgggcaagctgtttgtccccacaggtcctcggaatggcacagttcctcaggtgtc 362
          |||| ||||||||||| |||| ||||||||||| |||||||||| ||||||| |||||
Db   1535 GCTCCGGCAGCTGTTGTGCCTCCCACAGGTCCTTGGAATGGCACCAGTTCCTACAGGTGTC 1476

Qy    363 ttcacagatatcccaatcagcaacattcgtcgggttattgcacagcgatta 413
          |||||||||||||||||||||||||||||||| ||||     ||| |||
Db   1475 TTCACAGATATCCCAATCAGCAACATTCGTCGCGGTTAAGAGAATTACCATCA 1425
```

```
RESULT 13
AAS92326
ID   AAS92326 standard; cDNA; 2595 BP.
XX
AC   AAS92326;
XX
```

```
RESULT 14
AAA81500
ID   AAA81500 standard; DNA; 22863 BP.
```

```
XX
AC   AAA81500;
XX
DT   04-DEC-2000 (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_47 SEQ ID NO:47.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO20022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI   Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI   Rappuoli R, Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 1310-1317; 1760pp; English.
XX

CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neisseriae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 22863 BP; 5657 A; 5469 C; 5921 G; 5816 T; 0 other;
```

Query Match   8.8%;   Score 88.2;   DB 21;   Length 22863;
Best Local Similarity   46.6%;   Pred. No. 2.4e-14;
Matches 420;   Conservative 0;   Mismatches 448;   Indels 33;   Gaps 3;

```
Qy      85 ccgctgttcctccaacctcccagccttagctccataccttcagcaacctgccagcta 144
           |||||  ||  ||  ||| |||||  | || |||| ||| ||||| ||||| ||||||
Db    8216 cagctgctcactgcgcgctcgaagcccgcacgccgctcctgcagaagctgcccagctg 8275

Qy     145 ctcctgctgaccaaagggaagggtgtttgttagcctcttgcaaagaaagtggcagtag 204
           |||||  ||  ||   |   ||  ||   |||  ||| ||||  ||||  |  ||||||
Db    8276 ccgctcctgctgctacacaaaaacaacgcctatgcctgctgccatggcgcaactggctgcg 8335

Qy     205 agaaaggggattgatcttacacaagtaaaagggacaaggaaccagagttgaatcaaca 264
           || ||  ||  ||  || ||||  |   ||   || ||||  ||||  ||  |||  |
Db    8336 agaccggttgtgacgtgaacgcattgcaaggttccgtgccgcagtcgtcgtattgaaa 8395

Qy     265 aggatatcgaactcttttgtgcctagtaaagttgctcctccgcgcagctgttgtgtctc 324
           |  || |||  |  |||||| ||  ||||||  ||| ||||||| ||  ||  | ||||
Db    8396 aagacgtacaaaatcgccgtcgccaaacctgccggagccgcgtcgtcctgctgttgcactc 8455

Qy     325 ccaacaggtcctggaatggcacagttcctcacagtatccaatcggca 384
           ||  || ||| || ||   ||| |  |||| ||| ||||  || ||
Db    8456 ctgccggcgcacgtccctgaagaacgcg----------taccaatgagc 8494

Qy     385 acattcgttcgggttattgcacagcagcgatcaaatgcaatcaaaagcaaaacatacctccattatt 444
           | || |||| ||| ||||||| | ||| ||||  || ||| |||||||| | ||||| |||
Db    8495 gcctgcggctcccgtgttgcagaacgccttcttggcttccaacaagaaaacgccattctga 8554

Qy     445 aacttctatcgtatcgtaataatatgggaagaagtttcgttggtacgggaaagaactaataaga 504
           || | |||||| |||| | ||||||| | ||  |||||||  || | |||| ||| |||||  |
Db    8555 ctacattcaacgaggtcaacatcggatcatccgaaaccaatcatggactgctgcggttggtgtgccaa 8614

Qy     505 tatt-----agaagggagagaagcaaaaatttctgtcaatgactcatcataaaagcttag 558
           ||||      | || || |  |  ||||  | | | |  ||| |||| |||| ||| |||
Db    8615 aattcagagaaagaacacgcgtgaaactgggcttattgtcctcttcgttaaagcgctg 8674

Qy     559 cttcggcatgtttaaaagttcccgaaagcaaaattcttcttggaatggaccaagttataagac 618
           |||  |||||  |  |||||||| || | |||| || ||  ||| |||  |||| |||  |
Db    8675 ttgccgcccctgaaaataccggtcttgaatgtcttcgttgacggacaaagatcgtgt 8734

Qy     619 aaaatcatgttgttgatgtcagtgtcgttgcggttcgtctgctgtacctccgcaggactgtcacacacta 678
           | |||||||||  |  |||  ||| ||||| ||||  |||| ||| ||||| ||||  ||||  ||
Db    8735 accacggctacttcgacatcggtatccaattggcagccacgcggttggtgtccgca 8794

Qy     679 ttgtcgtttaatgcacatattaaaagagggagaaacccattgctaatgatgttgttcttag 738
           ||||| ||| ||||  | | |||  || ||||  ||| | ||| | |||| || || |||
Db    8795 ttctgcgctgatgcgaccaaaagcattgccgacatcgaacacaagcaattgttgattacg 8854

Qy     739 caaccaaagcaagagagggtaaactacagcacacatgaattccaaggtggcacttttacga 798
           || ||  |||| |  ||||  ||| | ||||| | |||| ||||  ||| || ||| |||
Db    8855 cgaaaaaaggccaaagacgcaaagatcgtcatcgaagatctgaccggggtacattcagta 8914

Qy     799 tctccaattaggagaatgttgttggattaagaaattctctgctattatttaaccaccctcaag 858
           |||  |||  |||| |||||| || || || |||| ||| ||| | |||| |||||| ||||
Db    8915 ttacca.cggcgggtacttcggttctatgatgtctatccgatcatcaaccacctccaat 8974

Qy     859 catgtatttggcaattggtgcttcagaggataaactgtccccgcagataatgaaaag 918
           ||| | |||| |  |||  ||| |||||| |  ||| || |||  ||| | |||||
Db    8975 ctgcgatttggtattgcacgcccac-----taaagagagcgcggggttgaaaacggc 9028

Qy     919 ggttgatggctagcatgatgtctgttacactcagttgtgtatcacgcggtggtcgatg 978
           ||||| ||| |  ||||||||  ||| | |||||| || ||||  ||  |||| |||
Db    9029 aagtgttgttccgatgatgtatctcgtctt acgacacacggtatatcattgacg 9088

Qy     979 g 979
           |
Db    9089 g 9089
```

KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US05928.
XX
PR   30-APR-1999;   99US-0132068.
PR   08-OCT-1999;   99WO-US23573.
PR   28-FEB-2000;   2000GB-0004695.
XX
PA   (CHIR) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI   Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI   Frazer CM, Grandi G;
XX
WPI: 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -
XX
PS   Claim 7; Appendix A; 692pp; English.
XX

CC   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own it was split into 8
CC   sequences which overlap each other at the beginning and end of each
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   the present invention. The NMB genome and fragments from it have
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   proteins can be used in compositions for treating or preventing infection
CC   due to Neisseria bacteria or as a diagnostic reagent for detecting the
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   bacteria. Computers, computer memory, computer storage medium or computer
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   more effective in vaccines than the outer membrane proteins currently
CC   used.
XX
SQ   Sequence 349980 BP; 81351 A; 86755 C; 95584 G; 86290 T; 0 other;

Query Match   8.8%; Score 88.2; DB 21; Length 349980;
Best Local Similarity   46.6%;
Matches 420; Conservative   0; Mismatches 448; Indels 33; Gaps   3;

```
Qy        85 ccgctgtccctccaaactcccagcgcctttagctcctacacacctgccagcta 144
              |  || ||| |||||  ||| || | ||| |  || ||| |||||||||
Db     70787 cagtgtgctactgccgctgctgaagccccagccgcgcctcctgcagaagagtgcccagctg 70846

Qy       145 ctcctgctggaccaaaggggaagggtgttcgttagccctctcttgcaaagaaagttggcagtag 204
              ||||||| ||| |||| | ||||  |||| ||  || ||||||
Db     70847 ccgctcctgctgctacacaaaacacgccgctatgcctgctgccgccaaactggctgccg 70906

Qy       205 agaaagggattgatcttacacaagtaaaaggaccagagacgatagttagaatcaccagga 264
              ||| |  |||| || ||| | |||   |||| |||   |
Db     70907 agaccggttgactgaacgcattgcaaggttccgccgtgacggtcgcgtattgaaag 70966

Qy       265 aggatatcgaactctttgtgcctagtaaagttgctcctcggcagtgtgtgtgcctc 324
              |  || || | |  ||| |  ||| |||| ||  |||| |||| || | ||
Db     70967 aagacgtacaaaatgcgctgcaaactgccggacggcgctgctcctgcttgcacttc 71026

Qy       325 ccacaggtcctggaatggcaccagttccttacacaggtgtgtcttcacagatatcccaaatcagca 384
              ||| | ||| |  |||| | ||  |  | | |||||||| ||| | ||    | ||
Db     71027 ctgccggccgcacgtccctgaagaacgcg-------------------taccaatgagcc 71065

Qy       385 acattcgttcgggttattgcacacgtgcacacctgccagcta 444
              || ||||| |  ||| | || |  |  |||||| ||||||||
Db     71066 gcctgcgtgccccgtcgttgcagaacgccttggcttccaacaagaaaacgccattctga 71125

Qy       445 accttctcatcgatgtaaatatgggagaagttctgttggtacggaaagaagaacttaataga 504
              |||||| | || ||| || || ||||| | ||||| ||| || |   | || |
Db     71126 ctacattcaaccgaagtcaacatggaacccaatcatcggacttgctgtgcgaagtacaaagaa 71185

Qy       505 tatt------agaaggagaagcaaaatttctgtccaatgactcatccatcaaaagcttcag 558
              |||       ||||||||| ||  | | |||||| | |||||||| || | ||| | |
Db     71186 aattccagaagaacacggccgtgaaactcgggcttaatgtcctctcttcgttaaaagccgctg 71245

Qy       559 cttggcatgttaaaagttccccgaaagcaaattcttctttgggatggaacacagttataagac 618
              || ||||| |||||  | ||| || | ||||||| |  | |||||| ||  | |  |||
Db     71246 ttgccgcccctgaaaaaatacccggttcgtgaatgcttctgcttgttgacggcaaagcatcgt 71305

Qy       619 aaaatcaatgttgttcgatgtcagttgcgtcagtactcctgcaggactcatcacacacta 678
              || |  ||| |||| ||| || | | |   |  |  |||| | ||| | ||  |||| |
Db     71306 accacggctacttcgaacatcggtatccgcaaattcgcagcgcccacgcggttcggttgtcgcaa 71365

Qy       679 ttgtgtttaatgcacatataaaggagttggaaaccattgctaatgatgtttgttcttag 738
              ||||| | || ||| |  |||  ||| ||||| | || || ||| ||||| |||  ||
Db     71366 ttctgcgtgatgccgaccaaatgagcattgccgacacagcaaattgttgattacg 71425

Qy       739 caaccaaaagcaagagagggtaaactacagccacatgaattccaaggtggcacttttacga 798
              |  || |||  ||||  || ||| | ||||||| |  || |||||||| ||  || | |
Db     71426 cgaaaaaagcccaaagacggcaaaatcgtatcgacagatccgcggcgtgtacattcagta 71485

Qy       799 tctccaattaggaatgttggattaagaattctctgctattattaaccacctcaag 858
              | ||| | ||||| |||| || | || | | || ||| ||| ||||| || |||
Db     71486 ttaccaacgcggtacttcggttcatgtctacccgatcatcaaccacactcaat 71545

Qy       859 catgtattttggcaattgtgctctcgttcagaggataaactggtccctgcagtaatgaaaag 918
              ||||  |||||| ||||| |  |  |   || | | || | || |  || || || |   || |
Db     71546 ctgcggattttgggtatgcacgccac--------taaagagcgcggtcggttgaaaaacggc 71599

Qy       919 ggtttgatgtggctagcatgagtgtctgttacactcagttgtgatcacccgggtggtggatg 978
              | ||||| ||||| |||  ||  | |  |||| | ||||| ||  ||||  |||| ||| ||
Db     71600 aagttgttccgtccgatgatgtatctgctcttacgaccacaccgtatcattgacg 71659

Qy       979 g 979
              |
Db     71660 g 71660
```

Search completed: August 13, 2002, 19:38:36
Job time: 4333 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: August 13, 2002, 17:37:48 ; Search time 49.87 Seconds
                                                    (without alignments)
                                                    4930.402 Million cell updates/sec

Title: US-09-761-580-1_COPY_800_1800
Perfect score: 1001
Sequence: 1 atatatcagcattgctgac...........cagtggagcccagtggctt 1001
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Database :   Issued_Patents_NA:*
                       1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                       2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                       3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                       4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                       5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                       6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

_al number of hits satisfying chosen parameters: 767066

searched: 383533 seqs, 122816752 residues

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	85	8.5	1708	3	US-09-108-020-5	Sequence 5, Appl1
2	69.8	7.0	2706	4	US-09-066-046-23	Sequence 23, Appl1
3	69.8	7.0	3073	4	US-08-975-762-41	Sequence 41, Appl1
4	69.8	7.0	3073	4	US-09-295-028-41	Sequence 41, Appl1
5	69.8	7.0	3073	4	US-09-106-582-41	Sequence 41, Appl1
6	65.2	6.5	4623	1	US-08-370-193A-7	Sequence 7, Appl1
7	62.2	6.2	1039	4	US-08-975-762-36	Sequence 36, Appl1
8	62.2	6.2	1039	4	US-09-821-324-36	Sequence 36, Appl1
9	62.2	6.2	1039	4	US-09-295-028-36	Sequence 36, Appl1
10	62.2	6.2	1039	4	US-09-106-582-36	Sequence 36, Appl1
11	48.6	4.9	1096	3	US-08-881-771A-5	Sequence 5, Appl1
12	48.4	4.8	426	3	US-08-881-771A-8	Sequence 8, Appl1
13	48.4	4.8	2793	1	US-08-458-298-1	Sequence 1, Appl1
14	42.6	4.3	2793	3	US-09-108-020-15	Sequence 15, Appl1
15	38.6	3.9	1450	3	US-09-050-863-2	Sequence 2, Appl1
16	38	3.8	2580	3	US-09-359-081-2	Sequence 2, Appl1
17	38	3.8	2580	4	US-08-975-762-36	Sequence 36, Appl1
18	38	3.8	5452	2	US-09-130-114-1	Sequence 1, Appl1
19	38	3.8	9600	4	US-08-910-647-1	Sequence 1, Appl1
20	38	3.8	10596	1	US-07-884-811-15	Sequence 15, Appl1
21	38	3.8	10596	1	US-07-885-971-15	Sequence 15, Appl1
22	38	3.8	10596	1	US-08-087-783A-15	Sequence 15, Appl1
23	38	3.8	10596	2	US-08-194-088B-15	Sequence 15, Appl1
24	38	3.8	10596	1	US-08-194-087-15	Sequence 15, Appl1
25	38	3.8	10596	5	PCT-US93-04648-15	Sequence 15, Appl1
26	37	3.7	913	1	US-08-217-327-3	Sequence 3, Appl1
27	37	3.7	913	1	US-07-885-970A-3	Sequence 3, Appl1
28	37	3.7	913	1	US-08-298-687A-3	Sequence 3, Appl1
29	37	3.7	913	1	US-08-530-797-2	Sequence 2, Appl1
30	37	3.7	913	1	US-08-298-829-3	Sequence 3, Appl1
31	37	3.7	913	1	US-08-787-335-2	Sequence 2, Appl1
32	37	3.7	1984	1	US-07-885-970A-25	Sequence 25, Appl1
33	37	3.7	1985	1	US-08-298-687A-25	Sequence 25, Appl1
34	37	3.7	1985	1	US-08-298-829-25	Sequence 25, Appl1
35	36.8	3.7	934	2	US-08-731-722-6	Sequence 6, Appl1
36	36.8	3.7	1505	1	US-07-915-246-15	Sequence 15, Appl1
37	36.6	3.7	1186	2	US-08-731-722-5	Sequence 5, Appl1
38	36.2	3.6	1748	3	US-09-100-730-1	Sequence 1, Appl1
39	36	3.6	11299	4	US-09-238-356-14	Sequence 14, Appl
40	35.6	3.6	342	1	US-09-056-556-219	Sequence 219, Appl
41	35.6	3.6	515	4	US-09-056-556-220	Sequence 220, App
42	35.6	3.6	16442	3	US-08-781-891-208	Sequence 208, App
43	35.2	3.5	2338	1	US-08-425-069-1	Sequence 1, Appl
44	35.2	3.5	2338	1	US-08-317-844B-1	Sequence 1, Appl1
45	34.2	3.4	578	6	5242821-15	Patent No. 5242821

ALIGNMENTS

RESULT 1
US-09-108-020-5
Sequence 5, Application US/09108020A
Patent No. 6143561
; GENERAL INFORMATION:
;   APPLICANT: Randall, Douglas D.
;   APPLICANT: Johnston, Mark L.
;   APPLICANT: Miernyk, Jan A.
;   APPLICANT: Luethy, Michael H.
;   APPLICANT: Mooney, Brian P.
;   TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
;   TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
;   TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANOATE BIOSYNTHESIS IN PLANTS
;   FILE REFERENCE: UMO 1482
;   CURRENT APPLICATION NUMBER: US/09/108,020A
;   CURRENT FILING DATE: 1998-06-30
;   EARLIER APPLICATION NUMBER: 60/076,544
;   EARLIER FILING DATE: 1998-03-02
;   EARLIER APPLICATION NUMBER: 60/051,291
;   EARLIER FILING DATE: 1997-06-30
;   EARLIER APPLICATION NUMBER: 60/055,255
;   EARLIER FILING DATE: 1997-08-01
;   NUMBER OF SEQ ID NOS: 54
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1708
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-108-020-5

Query Match   8.5%; Score 85; DB 3; Length 1708;
Best Local Similarity   52.8%; Pred. No. 1.3e-15;
Matches 257; Conservative   0; Mismatches 220; Indels 10; Gaps 3;

Qy   505 tatagaagggagagaagcaaaattctcgtaatgagttcatcataa----aagcttcagct 560
          || | |||||| ||| |||| | |||||||||||||| ||| || ||||| |||||
Db   953 tacgagaaggtgaagcaaagggtgtaacacatgacagcttaattagctaaagctgcaggg 1012

Qy   561 ttggcatgttaaaagttcccgagcgcaaattcttcttggatggac---acagttataga 617
          ||||| |||| |||| | | |||||| |||||||| | || ||||| |||| |
Db   1013 atggccttggctcagcatcctgtgttgaacgctagctgcaaagacgggaggagtttagt 1072

Qy   618 caaaatcatgttgtttgatgtcagttgcggtcggtcagtactcctgcaggactcatcacact 677
          || | ||| |||| |||| || | ||| ||| |||| | | ||| || ||| | ||||
Db   1073 tacaatagtagcattaacattgcagtggccggttgctatcaatgtggcctgattactacgct 1132

Qy   678 attgtgtttaatgcacatataaaagggagtggaaaccatgctaatgatgttgttcttta 737
          |||| ||| | || | | | |||| || || ||||| ||| | |||| |||||
Db   1133 gttctacaagatgcagataagtcagagattgttgttgtactcgtgtatctcaaaaatggaaaagagctg 1192

```
Qy    738  gcaacaaagcaagagaggggtaaactaacagccacatgaatccagggtgcacttttacg  797
           ||||| |||  ||||||  ||| ||||||||||||  |||||| |||||||||||||||
Db   1193  gtggggaaagctagaagcaactcaacccatgaatacacaacctggaactttact 1252

Qy    798  atctccaattaaggaatgtttgggaattaagaaattttctctgtctattattaaccacctcaa  857
           ||||| || || |||| ||| |||||||||||||||||| ||||||||||||||||| |||||
Db   1253  ttatcgaatccggtatgtttggagtggatagagttgacgctattcttccgccaggacag 1312

Qy    858  gcatgtgtatttgggtcttcaggaggataaactggtcccctgccagataataatgaaaa  917
           ||| ||| ||||| ||||||||||||| ||||| ||  |||  || |||| ||||||||
Db   1313  ggtgctattaggctgttgggagtcaaagccaactgtagt---tgctgataaggatgga 1369

Qy    918  gggttgatgtggctagacatgatgtctgttacactcaagttgtgatcaccgggtggtggat  977
           |||||| ||||||| |  | ||||| ||| |  |||||| ||||| |||||| ||||| |||
Db   1370  ttcttcagtgtaaaaacacaatgtcggtgactgagatcatcgcaattgtgtat 1429

Qy    978  gggacag  984
           ||||| ||
Db   1430  gggagctg 1436
```

..SULT   2
US-09-066-046-23
; Sequence 23, Application US/09066046A
: Patent No. 6204252
GENERAL INFORMATION:
    APPLICANT: MURPHY, Cheryl
              STOREY, James
              BELTZ, Gerald A.
              COUGHLIN, Richard T.
    TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC
                       EHRLICHIA AND METHODS OF USE
    NUMBER OF SEQUENCES: 65
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: HALE AND DORR LLP
        STREET: 60 State Street
        CITY: Boston
        STATE: MA
        COUNTRY: United States
        ZIP: 02109
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/066,046A
        FILING DATE: 24-Apr-1998
        CLASSIFICATION: <Unknown>
    ATTORNEY/AGENT INFORMATION:
        NAME: Superko, Colleen
        REGISTRATION NUMBER: 39,850
        REFERENCE/DOCKET NUMBER: 106.941.155
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (617) 526-6000
        TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 23:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 2706 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    HYPOTHETICAL: NO
    ANTI-SENSE: NO
    SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-066-046-23

Query Match   7.0%;   Score 69.8; DB 4;   Length 2706;
Best Local Similarity 46.8%;   Pred. No. 6e-11;
Matches 258; Conservative 0; Mismatches 287;   Indels 6; Gaps   1;

```
Qy    378  atcagcaacattcgtcgggtttattgcacagcgattaatgcaatcaaagcaaaccatacct  437
           ||||| || || |||||||||||  ||  || ||||||| |||||| || ||| |||||||
Db    855  ATGAGCAAAATCCGCCAAGTATAGAGCTGCTTAGGCTTAAGGAGTTCACAAAAATACCTGCT 914

Qy    438  cattattaccttctctatcgatgtaaatatgggagaagttctgttttggtacgggaaagaactt  497
           |||||||| | ||| |  || |||| ||||| |||| ||||| |||||||| ||| |||| ||
Db    915  ACACTCAGCACACTTTAATGAAGTTGATTATGACGAAAGTGATGGAGCTCAGACTAAGTAC 974

Qy    498  aataagatattagaaggagg------aagcaaaattcctgtcaatgacttcatcataaaa  551
           |||| ||  || || |||     | ||| |||| |||  ||| ||||| |||| ||||||
Db    975  AAAGATGCCTTTGTTGAAGAGGGTATGAGTGTTAAGCTTGGGTTATGCTTCTTTTATCAGA 1034

Qy    552  gcttcaagctttggcatgttaaaagttcccgaagcaaattctttcttggatggacaagtt  611
           |||| |||  || ||| ||||| ||||  |  || |||||| ||| |||| ||||||||| |
Db   1035  GCGGTTGTGCCTAGTCCTTTCCGAAAATTCCTGTCGTCGTGTCCGAATGCCGAGATTCAGGCGGATGAT 1094

Qy    612  ataagacaaaatcatgtgttgatgtcagtgttgcgtgctactcctgcgaggactcatc  671
           || | || |||| |||| ||| ||| |||| || |  ||| || ||| ||| |||| |
Db   1095  ATAGTCTACAGGGACTATTGTAACATTGGACTGCCGGTACCGATAAGGGGTTAGTG 1154

Qy    672  acacctattgtgttaaatgcaacatataaaaggagtggaaaccattgctaatgttgtt  731
           |||  || | |||||| || |||||| || ||| |||||||| || ||||| |||||
Db   1155  GTGCCTGTTATCAGAAGAGAGCGGGAAACTATTGTCACTTGCTGTGAAATGGAGCAACGACACTTGTT 1214

Qy    732  tccttagcaaccaaagcaagagaggggtaaactacaagccacatgaattccaaggtggcact  791
           ||||| || |||| ||||| ||||  || ||||  |||| ||| || ||||||||| ||||
Db   1215  GACTTAAGTACAAAAAGCAAGAAGAGTGGCAAGCTCTTCGTTTCGATATGTTCTGGTGCAACC 1274

Qy    792  tttacgatctccaatttaggaagtgtttggaattaagaatttctctgtctattattaaccca  851
           ||| || ||||| ||||||| ||| || ||| || |  |||||||| ||||||||||| ||
Db   1275  TTTACTATTACCAATGGTGGTCGTGTTATGGGTCGCGTCGTATTGTGTCTACCCTATAATCAAACCCT 1334

Qy    852  cctcaagcatgtattttggcaattggtgcttcagagggatcaaactggtccctgcagataat  911
           |||| || ||| || |||| |||| || ||| |  |||||  |||| ||||| || || |  |
Db   1335  CCTCAAATCTGGAAATCTTGGGTATGCATGCTATACAGCAGCACGTCCGTGCGTAGATGGT 1394

Qy    912  gaaaaagggtt  922
           ||| |||  |
Db   1395  AAGGTAGAGAT 1405
```

RESULT 3
US-08-975-762-41/c
; Sequence 41, Application US/08975762
: Patent No. 6207169
GENERAL INFORMATION:
    APPLICANT: Reed, Steven G.
    APPLICANT: Lodes, Michael J.
    APPLICANT: Houghton, Raymond
    TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
                       TREATM
    NUMBER OF SEQUENCES: 73
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: SEED and BERRY LLP
        STREET: 6300 Columbia Center, 701 Fifth Avenue
        CITY: Seattle
        STATE: Washington
        COUNTRY: USA
        ZIP: 98104
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/975,762
        FILING DATE: 21-MAR-1997
        CLASSIFICATION: 424
    ATTORNEY/AGENT INFORMATION:
        NAME: Maki, David J.
        REGISTRATION NUMBER: 31,392
        REFERENCE/DOCKET NUMBER: 210121.439
    TELECOMMUNICATION INFORMATION:

```
;   TELEPHONE: 206-622-4900
;   TELEFAX: 206-682-6031
;   INFORMATION FOR SEQ ID NO: 41:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 3073 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
US-08-975-762-41
```

```
Query Match          7.0%;  Score 69.8;  DB 4;  Length 3073;
Best Local Similarity  46.8%;  Pred. No. 6.5e-11;
Matches  258;  Conservative  0;  Mismatches 287;  Indels  6;  Gaps  1;

Qy   378 atcagcaacattcgtcgggttattgcacagcggattaatgcaatcaaagcaaaccatacct 437
             |||   | |||||| |||   | ||| | || ||| | || ||| || ||||||| |||
Db  1210 ATGAGCAAAAATCCGCCAAGTTATAGCTGCTAGGCTTAAGGAGTCACAAAAATACCTCTGCT 1151

Qy   438 cattattaccttctatcgatgtaaatatgggagaagtttttgttggtacggaaagaaactt 497
          || ||| || ||||| |  ||   |  || ||  || ||||| ||||| ||  |||  || ||
Db  1150 ACACTCAGCACCTTTAATGAAGTTGATATGACCAAAGTGACTTCAGACCTAAGTAC 1091

Qy   498 aataagatattagaaggggag------aagcaaaaattctgtcaatgacttcatcataaaa 551
          || |||| | ||  ||  ||       ||||| | |||  | ||| | || ||| || |||
Db  1090 AAAGATGCCTTTGTGAAGAGGTATGAGAGTTAAGCTTGGGTTTATGTCCTTCTTTATCAGA 1031

Qy   552 gcttcagctttggcatgttttaaaagttcccgaagcaaatttcttcttggatggacacagtt 611
          |||  | || ||| | ||||||||| || || ||||| | |||| ||||| |||  |||||
Db  1030 GCCGTTGTGCTAGTCCTTTCCGAAATTCCTGTGCCGAATTCAGCCGATGAT 971

Qy   612 ataagacaaaatcatgttgttgatgtcagtgttgcggtcagtactcctgcaggactcatc 671
          || ||||| || ||| | ||| |  ||| || |  ||||  |||| ||| ||| | ||
Db   970 ATAGTCTACAGGGACTATTGTAACATTGGAGTCGCGCGTAGGTTACCGGAGATTTCAGGCGATGAT 911

Qy   672 acacctattgtgttaatgcacatataaaagaggagtggaaaccattgctaatgatgttgtt 731
             || ||| || | ||| ||||||||||| ||| | || ||| || || | |||| ||
Db   910 GTGCCTGTTATCAGAGAGAGAGGCGGAAACTATGTCACTTGCTGAAAATGGAGCAAGCACTTGTT 851

Qy   732 tctttagcaaccaaagcaagagagaggtaaactacagccaacatgaattccagggttggggcact 791
          |   ||| ||| || | ||| |||| || ||| || || | |||| |  |||||| ||||
Db   850 GACTTAAGTACAAAAGCAAGAAAGTGGCAAGCTCTCGTTTCTCGATATAGTCTCGGTGCAACC 791

Qy   792 tttacgatctccaatttaggaatgtttggcaattaagaattctctgctattattaaccca 851
          |  ||| ||| ||||| || ||||| || |||  |||| |  | ||| ||| |||||||||
Db   790 TTTACTATTACCAATGGTGTGTGTATGGGTATGCATCCTATACAGCAGACGGTCCTGTGTCGGCAGTAGATGGT 731

Qy   852 cctcaagcatgtatttggcttcagagagtaaactggctccctgcagataat 911
          |||||| ||| || |||| ||||| | | |||||  || |||||| || ||
Db   730 CCTCAACTGGAAACTTGGGTATGCATGCTATACAGCAGACGGTCCTGTGTCGGCAGTAGATGGT 671

Qy   912 gaaaaggggtt 922
          |  | ||| | |
Db   670 AAGGTAGAGAT 660
```

```
;   SEQ ID NO 41
;     LENGTH: 3073
;     TYPE: DNA
;     ORGANISM: Ehrlichia sp.
US-09-295-028-41
```

```
Query Match          7.0%;  Score 69.8;  DB 4;  Length 3073;
Best Local Similarity  46.8%;  Pred. No. 6.5e-11;
Matches  258;  Conservative  0;  Mismatches 287;  Indels  6;  Gaps  1;

Qy   378 atcagcaacattcgtcgggttattgcacagcggattaatgcaatcaaagcaaaccatacct 437
             |||   | |||||| |||   | ||| | || ||| | || ||| || ||||||| |||
Db  1210 ATGAGCAAAAATCCGCCAAGTTATAGCTGCTAGGCTTAAGGAGTCACAAAAATACCTCTGCT 1151

Qy   438 cattattaccttctatcgatgtaaatatgggagaagtttttgttggtacggaaagaaactt 497
          || ||| || ||||| |  ||   |  || ||  || ||||| ||||| ||  |||  || ||
Db  1150 ACACTCAGCACCTTTAATGAAGTTGATATGACCAAAGTGACTTCAGACCTAAGTAC 1091

Qy   498 aataagatattagaaggggag------aagcaaaaattctgtcaatgacttcatcataaaa 551
          || |||| | ||  ||  ||       ||||| | |||  | ||| | || ||| || |||
Db  1090 AAAGATGCCTTTGTGAAGAGGTATGAGAGTTAAGCTTGGGTTTATGTCCTTCTTTATCAGA 1031

Qy   552 gcttcagctttggcatgttttaaaagttcccgaagcaaatttctttcttggatggacacagtt 611
          |||  | || ||| | ||||||||| || || ||||| | |||| ||||| |||  |||||
Db  1030 GCCGTTGTGCTAGTCCTTTCCGAAATTCCTGTGCCGAATTCAGCCGATGAT 971

Qy   612 ataagacaaaatcatgttgttgatgtcagtgttgcggtcagtactcctgcaggactcatc 671
          || ||||| || ||| | ||| |  ||| || |  ||||  |||| ||| ||| | ||
Db   970 ATAGTCTACAGGGACTATTGTAACATTGGAGTCGCGCGTAGGTTACCGGAGATTTCAGGCGATGAT 911

Qy   672 acacctattgtgttaatgcacatataaaagaggagtggaaaccattgctaatgatgttgtt 731
             || ||| || | ||| ||||||||||| ||| | || ||| || || | |||| ||
Db   910 GTGCCTGTTATCAGAGAGAGAGGCGGAAACTATGTCACTTGCTGAAAATGGAGCAAGCACTTGTT 851

Qy   732 tctttagcaaccaaagcaagagagaggtaaactacagccaacatgaattccagggttggggcact 791
          |   ||| ||| || | ||| |||| || ||| || || | |||| |  |||||| ||||
Db   850 GACTTAAGTACAAAAGCAAGAAAGTGGCAAGCTCTCGTTTCTCGATATAGTCTCGGTGCAACC 791

Qy   792 tttacgatctccaatttaggaatgtttggcaattaagaattctctgctattattaaccca 851
          |  ||| ||| ||||| || ||||| || |||  |||| |  | ||| ||| |||||||||
Db   790 TTTACTATTACCAATGGTGTGTGTATGGGTATGCATCCTATACAGCAGACGGTCCTGTGTCGGCAGTAGATGGT 731

Qy   852 cctcaagcatgtatttggcttcagagagtaaactggctccctgcagataat 911
          |||||| ||| || |||| ||||| | | |||||  || |||||| || ||
Db   730 CCTCAACTGGAAACTTGGGTATGCATGCTATACAGCAGACGGTCCTGTGTCGGCAGTAGATGGT 671

Qy   912 gaaaaggggtt 922
          |  | ||| | |
Db   670 AAGGTAGAGAT 660
```

```
RESULT   4
US-09-295-028-41/c
;   Sequence 41, Application US/0929502B
;   Patent No. 6277381
;   GENERAL INFORMATION:
;     APPLICANT: Reed, Steven G.
;     APPLICANT: Lodes, Michael J.
;     APPLICANT: Houghton, Raymond L.
;     APPLICANT: McNeill, Patricia D.
;     TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;     TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
;     FILE REFERENCE: 210121.439C4
;     CURRENT APPLICATION NUMBER: US/09/295,028
;     CURRENT FILING DATE: 1999-04-20
;     NUMBER OF SEQ ID NOS: 85
;     SOFTWARE: FastSEQ for Windows Version 3.0
```

```
RESULT   5
US-09-106-582-41/c
;   Sequence 41, Application US/09106582
;   Patent No. 6306402
;   GENERAL INFORMATION:
;     APPLICANT: Reed, Steven G.
;     APPLICANT: Lodes, Michael J.
;     APPLICANT: Houghton, Raymond
;     TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
;     TITLE OF INVENTION: THERAP
;     NUMBER OF SEQUENCES: 73
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: SEED and BERRY LLP
;       STREET: 6300 Columbia Center, 701 Fifth Avenue
;       CITY: Seattle
;       STATE: Washington
;       COUNTRY: USA
;       ZIP: 98104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2

TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
TELEPHONE: 206-682-6031

INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-09-106-582-41

Query Match 7.08%; Score 69.8; DB 4; Length 3073;
Best Local Similarity 46.8%; Pred. No. 6.5e-11;
Matches 258; Conservative 0; Mismatches 287; Indels 6; Gaps 1;

```
Qy    378 atcagcaacattcgtcgggttattgcacagcgattaatgcaatcaaagcaaaccatacct 437
          |||||| ||  ||   |||||| |||||||| |||||||| |||||| ||| |  || |||
Db   1210 ATGAGCAAAAATCCGCCAAAGTTATAGCTTGCTAGGCTTAAGGAGTCACAAAATACCTCTGCT 1151

Qy    438 cattatcacttcctatcgatgtaaatatgggagagagttgttggtacggaaagcactt 497
          | |  ||  |  ||||||| || || || ||||  |  |||||||||| || ||| ||
Db   1150 ACACTCAGCACCTTTAATGAAGTTGATATGACGAAAGTAGAGTCAGAGCTAAGTAC 1091

Qy    498 aataagatattagaagggag------aagcaaaattctgcaaatgacttcatcataaaa 551
          ||| |  || | ||||| ||         || |||||| ||| || |  || || |||
Db   1090 AAAGATGCCTTTGTGAAGAGGTATGATGTTAAGCTTGGGTTTAGTGTCCTTCTTTATCAGA 1031

Qy    552 gcttcagctttggcatgttaaaagttcccgaagcaaattctcttggatggacacagtt 611
          | ||| ||||| |||| |||||||||  | || ||  |||| |||| |||| || |||
Db   1030 GCGGGTTGTGCTAGTCCTTTCCGAAAATTCCTGTGTCTGAATGCGGAGAGTTTCAGGCGATGAT 971

Qy    612 ataagacaaatcatgttgatgtcagttgcggtcagtactcctgcaggactcatc 671
          ||  || |||| ||||| |||| |  | |||| |||| |||| ||| || | ||
Db    970 ATAGTCACAGGGACTATTGTAACATTGGAGTCGCGGTAGGTTACCGATTAAGGGGTTAGTG 911

Qy    672 acacctattgtgttttaatgcacatataaaaggagtggaaaaccaattgctaatgatgttgtt 731
          || ||||| |  |||||||||| || ||||| ||| ||||||  || |||  ||||| ||||
Db    970 ATAGTCACAGGGACTATTGTAACATTGGAGTCGCGGTAGGTTACCGATTAAGGGGTTAGTG 911

Qy    732 tctttagcaaccaaagcaagagaggagggtaaactcacagccacatgaattccagggtggcact 791
          | ||| ||  |  || ||| |||| ||||  ||| |||||  |||| |||  || | || ||||
Db    910 GTGCCCTCGTTATCAGAGACGAGGACGGGAAACTATGTCACTTGCTGAAATGGAGCAAGCACTTGTT 851

Qy    850 GACTTAAGTACAAAAAGCAAGAAGAGTGGGCAAGCTCTCTGTTTCTGATATATGTCTCGGTGCAACC 791

Qy    792 tttacgatctccaatttaggaatgtttggaattaagaaattctctgctattattaaccca 851

Db    790 TTTACTATTACCAATGGTGGTGCGTGTATGGGTCGCTAGTGTCTACACCCTATAATCAACCCT 731

Qy    852 cctcaagcatgtgatttttggcaattggtgcttcagaggatataacctgtccctgcagataat 911
          |||||| | | |||||| |||||| | ||| || || || |  |||||| || |  ||||
Db    730 CCTCAATCTGGAATCTTGGGTATGCATGCTATACAGCAGCGCTCCTGTGGCACTAGAATGGT 671

Qy    912 gaaaaaggggtt 922
          ||| | |||  |
Db    670 AAGGTAGAGAT 660
```

APPLICANT: TSUJIMOTO, NOBUHARU
APPLICANT: MATSUI, KAZUHIKO
APPLICANT: KURAHASHI, KAZUHIKO
TITLE OF INVENTION: MUTANT AND METHOD FOR PRODUCING
TITLE OF INVENTION: L-GLUTAMIC ACID BY FERMENTATION
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
FILING DATE: 09-JAN-1995
APPLICATION NUMBER: US/08/370,193A
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-714-0

TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4623 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: 327..3128
OTHER INFORMATION: /note= "Method of feature
OTHER INFORMATION: determination: E"
FEATURE:
NAME/KEY: CDS
LOCATION: 3143..4357
OTHER INFORMATION: /note= "Method of feature
OTHER INFORMATION: determination: E"

US-08-370-193A-7

Query Match 6.5%; Score 65.2; DB 1; Length 4623;
Best Local Similarity 46.4%; Pred. No. 2e-09;
Matches 294; Conservative 0; Mismatches 328; Indels 12; Gaps 2;

```
Qy    352 ctacaggtgtcttcacagatatcccaatcagcaacattcgtcgggttattgcacagcgat 411
          |||| || | || ||||| |  |||| ||||||||||||||||||||| ||||| |||||
Db   3657 CTGCACGTAGTGAAAAACGTGTCCGATGACTCCCTGCGTAAGCGTGTGGCAGAGCGTC 3716

Qy    412 taatgcaatcaaagcaaaccatacctcattattactttcttctatcgatgtaaatatggggag 471
          ||| |||||| || |  |||| || |||| |||||  | || || |  |||||| ||||||
Db   3717 TGCTGGAACGGAAAAACTCCACCGCCATGACCACGTTCAACGAAGTCAACATGAAGC 3776

Qy    472 aagtttgttgtacggaaagcaacttaataagatattagaagggagaagcaaaatttct- 530
          ||| | ||| ||| || ||  |  ||||| ||| || ||  |||| || ||| |||||
Db   3777 CGATTATGGATCTCGGTAAGCAGTACGGTTGAAGCGGTTTGAAAAAACGGCACGGAGATCCGTC 3836

Qy    531 -----gtcaatgacttcatcataaaagcttcagctttggcatgttaaaagtcccgaag 585
               || |||| |||| ||| |||||| |||| ||||| |||| |||||||| |||||
Db   3837 TGGGCCTTTATGTCCTTCTACGTGAAGCGGTTGAAGCCCTGAAAACGTTACCCGGAAG 3896
```

RESULT 6
US-08-370-193A-7
Sequence 7, Application US/08370193A
Patent No. 5573845
GENERAL INFORMATION:
APPLICANT: ONO, EIJI

```
Qy  586 caaattcttcttggatggacacaagttataaagacaaaatcatgttgttgatgtcagtgttg 645
          |||||| ||| | |||| | |  ||  ||||| | ||  | | |||  ||| |||| || | |
Db 3897 TGAACGGCTTCTATCGACGGCGATGACGTGGTTTACCACACTATTTCGACGTTCAGCATGG 3956

Qy  646 cggtcagtactccctgcaggactcatcacacctattgtgttaatgcacatataaaggag 705
          |||| | || |||  || ||| ||  ||||||   | || ||||  | |||||| || |||
Db 3957 CGGTTTCTACGCGCCGCGGCCGGTTCGCGTCGATGTCGATACCCTCGGCA 4016

Qy  706 tggaaccattgctaatgatgttgttcttagcacacaaagcaaagagaggggtaaactac 765
          ||| | || |||| ||  || || |||| | |||| || | ||| ||| |||| || | ||
Db 4017 TGGCAGACATCGAGAAGAAAATCAAAGACCGTCCAGTCAACGGCAAGCTGA 4076

Qy  766 agccacatgaattccaggtgtggcacttcttttacgatctccaaatttaggaatgtttggaatta 825
          || || ||||| ||| |  ||||| ||||||  | ||| || |||| |  || |||| || | ||
Db 4077 CCGTTGAAGATCTGACGGTGGTTAACTTCACCATCACCAACGGTGGTGTTCGGTTCCC 4136

Qy  826 agaattctctgctattattaaccacctcaagcacatgtattttggcaattggtgttcag 885
          || ||||||| | ||| || ||| ||| |||| | || |||| ||| |||| || |||
Db 4137 TGATGTCTACGCGCCGATCATCAACCCGCGCCAGAGCGGAATTCGGGTATGCACGCTATCA 4196

Qy  886 aggataaactgtccctgcagtaataaaagggtttgatgtggctagcacgatgtctg 945
          || || | ||| | | ||   | |||||| ||| ||||| | || ||| |||| ||
Db 4197 AAGATCGTCCGATGGCGCGTGAATGGTCA------GGTTGAGATCCTGCCGATGATGTACC 4250

Qy  946 ttacactcaagttgtgatcaccgggtggttggatgg 979
          || |||| ||| || | ||| || || ||||| |||
Db 4251 TGGCGCTGTCCTACGATCACCGTCTCGATCGATGG 4284
```

RESULT 7
US-08-975-762-36
Sequence 36, Application US/08975762
Patent No. 6207169
GENERAL INFORMATION:
   APPLICANT: Reed, Steven G.
   APPLICANT: Lodes, Michael J.
   APPLICANT: Houghton, Raymond
   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMEN
   NUMBER OF SEQUENCES: 73
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: SEED and BERRY LLP
      STREET: 6300 Columbia Center, 701 Fifth Avenue
      CITY: Seattle
      STATE: Washington
      COUNTRY: USA
      ZIP: 98104
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/975,762
      FILING DATE: 21-MAR-1997
      CLASSIFICATION: 424
   ATTORNEY/AGENT INFORMATION:
      NAME: Maki, David J.
      REGISTRATION NUMBER: 31,392
      REFERENCE/DOCKET NUMBER: 210121.439
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: 206-622-4900
      TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 36:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 1039 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
   MOLECULE TYPE: DNA (genomic)
   ORIGINAL SOURCE:
      ORGANISM: Ehrlichia

Query Match   6.2%; Score 62.2; DB 4; Length 1039;
Best Local Similarity   47.8%; Pred. No. 6.4e-09;
Matches 181; Conservative   0; Mismatches 198; Indels   0; Gaps   0;

```
Qy  544 tcataaagcttcagcttgggcatgttaaaagtttcccgaagcaaattcttcttgggatgg 603
          | | ||||| |||| | ||| |||| ||||| |||||||||  ||||||||||| ||||||
Db    2 TAATCAGACGGGTGTGCTAGTCCTTTCCGAAATTCCTGTGCTGAAATGCGGAGATTTCAG 61

Qy  604 acacagttataagacaaaatcatgttgttgatgtcagtgttgcggtcagtactcctgcag 663
          | ||||||||| ||||||||||||||||||||||||||||| || |||| | ||||| |||
Db   62 GCGATGATATAGTCTACAGGGACTATTTGTTGAACATTGGGTCGCGGGTAGGTACGGATAAGG 121

Qy  664 gactcatcacacctattgtgttaatgcacatataaaggagtggaaaccattgctaatg 723
          |||| |||| ||| || || | |||| | | | |||| |  |||||| |||| | ||
Db  122 GGTTAGTGGTGCCCTGTTATCAGAAGAGCGGAAACTATGTCACTTGCTGAAATGGAGCAAG 181

Qy  724 atgttgttcttagcacacaaagcaaagagaggggtaaactacacgccacatgaattccagg 783
          ||| ||| || ||||| ||||| |||| || |||| |||| ||| || |||||||||| ||
Db  182 CACTTGTTGACTTAAGTACAAAAAGCAAAGAAGTGGCAAGCTCTCTGTTTCTGATATGTCTG 241

Qy  784 gtggcacttcttacgatctccaaatttaggaatgtttggaattaagaaatttctctgctatta 843
          ||| ||| || ||  ||| || |||||| || || || || || | |||||||||| | ||
Db  242 GTGCAAACCTTTACTATTACCAAATGGTGGGTGTGTGATGGGGTCGCATTGTCTACCCCTATAA 301

Qy  844 ttaacccaccctcaagcatgtatttgtggtcgcttcagaggagaataactggtccctg 903
          || |||| ||| ||||||| | ||| ||||| ||| |||| || | || |||| || |
Db  302 TCAACCCTCCTCAAACCTGGAATCTTGGGTATGCATGCTATACACAGCACGCGTCCTGTGGCAG 361

Qy  904 cagataataatgaaaaggggtt 922
          |||| || || || |||||||
Db  362 TAGATGGTAAGGTAGAGAT 380
```

RESULT 8
US-08-821-324-36
Sequence 36, Application US/08821324
Patent No. 6231869
GENERAL INFORMATION:
   APPLICANT: Reed, Steven G.
   APPLICANT: Lodes, Michael J.
   APPLICANT: Houghton, Raymond
   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND THERAP
   NUMBER OF SEQUENCES: 38
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: SEED and BERRY LLP
      STREET: 6300 Columbia Center, 701 Fifth Avenue
      CITY: Seattle
      STATE: Washington
      COUNTRY: USA
      ZIP: 98104
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/821,324
      FILING DATE: 21-MAR-1997
      CLASSIFICATION: 424
   ATTORNEY/AGENT INFORMATION:
      NAME: Maki, David J.
      REGISTRATION NUMBER: 31,392
      REFERENCE/DOCKET NUMBER: 210121.439
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: 206-622-4900
      TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 36:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 1039 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-08-821-324-36

Query Match   Best Local Similarity   47.8%;   6.2%;   Score 62.2;   DB 4;   Length 1039;
Matches 181; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```
Qy  544 tcataaaagcttcagcttggcatgttaaaagttcccgaagcaaattcttcttggatgg 603
Db    2 TAATCAGAGCGGTTGTGTCCTAGTCCTTTCCGAAATTCCTGTGCTGGAGATTCAG 61

Qy  604 acacagttataagacaaaatcatgtgttgtttgatgtcagtgttgccggtcagtactcctgcag 663
Db   62 GCGATGATATAGTCTACAGGGACTATTGTATTAACATTGAGTCTCCCGGTAGGTACCGATAAGG 121

Qy  664 gactcatcacacctattgtgttttaatgcacatataaaaggagtggaaaccattgctaatg 723
Db  122 GGTTAGTTGGTGTCCCTGTTTATCAGAACTACAGCCACATGAATTCCAGG 181

Qy  724 atgttgtttcttagcaaccaagcaagagaggagggtaaactacagcccaatgaattccag 783
Db  182 CACTTGTTGACTTAAGTACAAAAAGCAAGAAGTGCGAAGCTCTCGTTCGTTCTGATAATGTCTG 241

Qy  784 gtggcactttacgatctccaattaggaaatgtttggaattaagaattctctgctatta 843
Db  242 GTTCAACCTTTACTATATTACCAATGGTGTGGTCGGTCGGTCTATTGTTCTACCCTATPA 301

Qy  844 ttaaccacctcaagcatgtatttgcaatgtggttgcttcagaggataaaactggtccctg 903
Db  302 TCAACCCTTCCTCAAATCTGGAAATCTTGGGTATGTCATGCTGTATACAGCAGCGTCCTGTGCCAG 361

Qy  904 cagataatgaaaaagggtt 922
Db  362 TAGATGGTTAAGGTAGAGAT 380
```

RESULT 9
US-09-295-028-36
Sequence 36, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 36
LENGTH: 1039
TYPE: DNA
ORGANISM: Ehrlichia sp.
US-09-295-028-36

Query Match   Best Local Similarity   47.8%;   6.2%;   Score 62.2;   DB 4;   Length 1039;
Matches 181; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```
Qy  544 tcataaaagcttcagcttggcatgttaaaagttcccgaagcaaattcttcttggatgg 603
Db    2 taatcagagcggttgtgtccttccgaaatcctgtgctggagattcag 61

Qy  604 acacagttataagacaaaatcatgtgttgtttgatgtcagtgttgccggtcagtactcctgcag 663
Db   62 gcgatgatatagtctacagggactattgtataacattgagtctcccggtaggtaccgataagg 121

Qy  664 gactcatcacacctattgtgttttaatgcacatataaaaggagtggaaaccattgctaatg 723
Db  122 ggttagttggtgtccctgtttatcagaactacagccacatgaattccagg 181

Qy  724 atgttgtttcttagcaaccaagcaagagaggagggtaaactacagcccaatgaattccag 783
Db  182 cacttgttgacttaagtacaaaaagcaagaagtgcgaagctctcgttcgttctgataatgtctg 241

Qy  784 gtggcactttacgatctccaattaggaaatgtttggaattaagaattctctgctatta 843
Db  242 gttcaacctttactatattaccaatggtgtggtcggtcggtctattgttctaccctataa 301

Qy  844 ttaaccacctcaagcatgtatttgcaatgtggttgcttcagaggataaaactggtccctg 903
Db  302 tcaacccttcctcaaatctgggaaatcttgggtatgtcatgctgtatacagcagcgtcctgtgccag 361

Qy  904 cagataatgaaaaagggtt 922
Db  362 tagatggttaaggtagagat 380
```

RESULT 10
US-09-106-582-36
Sequence 36, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND THERAP
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 1039 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Ehrlichia
US-09-106-582-36

Query Match   Best Local Similarity   47.8%;   6.2%;   Score 62.2;   DB 4;   Length 1039;
Matches 181; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```
Qy  664 gactcatcacacctattgtgttttaatgcacatataaaaggagtggaaaccattgctaatg 723
Db  122 ggttagttggtgtccctgtttatcagaactacagccacatgaattccagg 181

Qy  724 atgttgtttcttagcaaccaagcaagagaggagggtaaactacagcccaatgaattccag 783
Db  182 cacttgttgacttaagtacaaaaagcaagaagtggcaagctctcgttcgttctgataatgtctg 241

Qy  784 gtggcactttacgatctccaattaggaaatgtttggaattaagaattctctgctatta 843
Db  242 gttcaacctttactatattaccaatggtgtggtcggtcggtctattgttctaccctataa 301

Qy  844 ttaaccacctcaagcatgtatttgcaatgtggttgcttcagaggataaaactggtccctg 903
Db  302 tcaaccctctcctcaaatctgggaaatcttgggtatgtcatgctgtatacagcagcgtcctgtgccag 361

Qy  904 cagataatgaaaaagggtt 922
Db  362 tagatggttaaggtagagat 380

Qy  544 tcataaaagcttcagcttggcatgttaaaagttcccgaagcaaattcttcttggatgg 603
```

```
Db     2 TAATCAGAGCGGTGTTGCTACTTCCTTTCCGAAATTCCTGTGCTGCTGAATGCGGAGATTTCAG 61

Qy   604 acacagttataagacaaaatcaatgttgttgatgtcagtgttgcggtcagtactcctgcag 663
         |||||||| ||||||| ||||| ||||||||||||||||||||||
Db    62 GCGGATGATATAGTCTACAGGGACTATTGTTAACATTGGAGTGTGCCGGTTAGGTACCGATAAGG 121

Qy   664 gacttcatcaacctattgtgttaatgcacatataaaagagagtggaaaccattgctaatg 723
         ||||| | ||||| ||| ||||| ||||||||||||||||||||
Db   122 GGTTAGTGGTGCCCTGTTATCAGAAGAGAGAGCGGAAACATATGTCACTTGCTGAAATGGAGCAAG 181

Qy   724 atgtgtttcttagcaaaccaaagagaagggtaaactacagccacatgaattccagg 783
         ||||| ||||||||| ||| |||
Db   182 CACTTGTTGACTTAAGCTAACAAAAGCAAGAAGCTGGCAAGCTCTCTGTTTCTGATATCTCTG 241

Qy   784 gtggcacttacgtctccaattaggaattgtttggaattaagaatttctctgtatta 843
         ||||| ||||| |||
Db   242 GTGCAACCTTTACTATTACCAAATGGTGTGTGTATGGGTCGCTATTGTCTTACACCCTATTA 301

...  844 ttaaccccaccccaagacatgtatttggcaattggtgcttcagaggggtataaactggtccctg 903
         |||| |||||| ||| ||||||||||
Db   302 TCAACCCTCCTCCAAATCTGGAATCTTGGGTATGCATCCTATACAGCACAGCGTCCTGTGTGGCAG 361

Qy   904 cagataatgaaaagggtt 922
         |||| ||| |||||
Db   362 TAGATGGTAAGGTAGAGAT 380
```

RESULT 11
US-08-881-771A-5

; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-881-771A-5:

Query Match   4.9%;   Score 48.6;   DB 3;   Length 1096;
Best Local Similarity 85.7%;   Pred. No. 7.8e-05;
Matches   54; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

```
Qy     1 atatatcagcatttgctgacttataggccaaccgaagtaacagagatttaaaaccacaagtgc 60
         |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   755 ATATATCAGCATTTGCTGACTATAGGCCAAACCGAAGTAACACAGATTTAGAATTCCCGGTC 814

Qy    61 cac 63
         ||
Db   815 GAC 817
```

; Sequence 5, Application US/08881771A
; Patent No. 6111071
; GENERAL INFORMATION:
;   APPLICANT: Gershwin, Eric
;   APPLICANT: Leung, Patrick
;   APPLICANT: Coppel, Ross
;   TITLE OF INVENTION: RECOMBINANT FUSION PROTEIN
;   COMPRISING PDC-
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Mandel & Adriano
;   STREET: 35 N. Arroyo Parkway, Suite 60
;   CITY: Pasadena
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 91103
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/881,771A
;   FILING DATE: 24-JUN-1996
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Adriano, Sarah B
;   REGISTRATION NUMBER: 34,470
;   REFERENCE/DOCKET NUMBER: 30448.51US01
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 626-395-7801
;   TELEFAX: 626-395-0694
;   TELEX:
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1096 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO

RESULT 12
US-08-881-771A-8

; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-881-771A-8:

Query Match   4.8%;   Score 48.4;   DB 3;   Length 426;
Best Local Similarity 98.0%;   Pred. No. 5.1e-05;
Matches   49; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

; Sequence 8, Application US/08881771A
; Patent No. 6111071
; GENERAL INFORMATION:
;   APPLICANT: Gershwin, Eric
;   APPLICANT: Leung, Patrick
;   APPLICANT: Coppel, Ross
;   TITLE OF INVENTION: RECOMBINANT FUSION PROTEIN
;   COMPRISING PDC-
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Mandel & Adriano
;   STREET: 35 N. Arroyo Parkway, Suite 60
;   CITY: Pasadena
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 91103
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/881,771A
;   FILING DATE: 24-JUN-1996
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Adriano, Sarah B
;   REGISTRATION NUMBER: 34,470
;   REFERENCE/DOCKET NUMBER: 30448.51US01
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 626-395-7801
;   TELEFAX: 626-395-0694
;   TELEX:
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 426 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:
US-08-881-771A-8:

Qy   1 atatatcagcattgctgactataggccaaccgaagtaacaagatttaaaa 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374 ATATATCAGCATTGCTGACTATAGGCCAACCGAGTAACAGATTTAGAA 423

RESULT 13
US-08-209-747-1/c
; Sequence 1, Application US/08209747
; Patent No. 5733771
; GENERAL INFORMATION:
;   APPLICANT: Lewis, Randolph V.
;   APPLICANT: Colgin, Mark
;   TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
;   TITLE OF INVENTION: Silk Proteins
;   NUMBER OF SEQUENCES: 56
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Birch, Stewart, Kolasch & Birch
;   STREET: P.O. Box 747
;   CITY: Falls Church
;   STATE: Virginia
;   COUNTRY: USA
;   ZIP: 22040-3487
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/209,747
;   FILING DATE: 14-MAR-1994
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Murphy Jr., Gerald M.
;   REGISTRATION NUMBER: 28,977
;   REFERENCE/DOCKET NUMBER: 1447-104P
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 703-205-8000
;   TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2793 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Nephila clavipes
;   TISSUE TYPE: minor ampullate gland
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 183..2675
;   OTHER INFORMATION: /product= "N. clavipes minor
;   OTHER INFORMATION: ampullate silk protein"
US-08-209-747-1

Query Match   4.38; Score 42.6; DB 1; Length 2793;
Best Local Similarity 59.5%; Pred. No. 0.0085;
Matches   72; Conservative   0; Mismatches 49; Indels   0; Gaps   0;

Qy   38 aacagatttaaaacaacaagtgccacacctacccacaccccggtggccgcgctgttcctcc 97
          || ||| |||| | ||||| | |||||||| ||||||||||||||||||| ||| |||||
Db   1663 ACCATATTCAGCACCACCGGTGCCTGCACCTGCCACCTGCCCAGCGGCGCCACCAGCTCCAGCACC 1604

Qy   98 aactcccagcctttagctcctacacttcagcaccctgccccagctactcctgctggacc 157
          ||| |||||||| ||||| |||||||||||||||||||||||||| |||||| |||
Db   1603 AGCACCTCTACCGTAACCACCAGCCGCCTCCGGCTCCTGTTGCAGCAGCTGCAGCAGCACC 1544

Qy   158 a 158
          |
Db   1543 A 1543

RESULT 14
US-08-458-298-1/c
; Sequence 1, Application US/08458298
; Patent No. 5756677
; GENERAL INFORMATION:
;   APPLICANT: Lewis, Randolph V.
;   APPLICANT: Colgin, Mark
;   TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
;   TITLE OF INVENTION: Silk Proteins
;   NUMBER OF SEQUENCES: 56
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Birch, Stewart, Kolasch & Birch
;   STREET: P.O. Box 747
;   CITY: Falls Church
;   STATE: Virginia
;   COUNTRY: USA
;   ZIP: 22040-3487
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/458,298
;   FILING DATE: 02-JUN-1995
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/209,747
;   FILING DATE: 14-MAR-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Murphy Jr., Gerald M.
;   REGISTRATION NUMBER: 28,977
;   REFERENCE/DOCKET NUMBER: 1447-104P
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 703-205-8000
;   TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2793 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Nephila clavipes
;   TISSUE TYPE: minor ampullate gland
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 183..2675
;   OTHER INFORMATION: /product= "N. clavipes minor
;   OTHER INFORMATION: ampullate silk protein"
US-08-458-298-1

Query Match   4.38; Score 42.6; DB 1; Length 2793;
Best Local Similarity 59.5%; Pred. No. 0.0085;
Matches   72; Conservative   0; Mismatches 49; Indels   0; Gaps   0;

Qy   38 aacagatttaaaacaacaagtgccacacctacccacaccccggtggccgcgctgttcctcc 97
          || ||| |||| | ||||| | |||||||| ||||||||||||||||||| ||| |||||
Db   1663 ACCATATTCAGCACCACCGGTGCCTGCACCTGCCACCTGCCCAGCGGCGCCACCAGCTCCAGCACC 1604

Qy   98 aactcccagcctttagctcctacacttcagcaccctgccccagctactcctgctggacc 157
          ||| |||||||| ||||| |||||||||||||||||||||||||| |||||| |||
Db   1603 AGCACCTCTACCGTAACCACCAGCCGCCTCCGGCTCCTGTTGCAGCAGCTGCAGCAGCACC 1544

Qy   158 a 158
          |
Db   1543 A 1543

RESULT 15
US-09-108-020-15
; Sequence 15, Application US/09108020A
; Patent No. 6143561
; GENERAL INFORMATION:
; APPLICANT: Randall, Douglas D.
; APPLICANT: Johnston, Mark L.
; APPLICANT: Miernyk, Jan A.
; APPLICANT: Luethy, Michael H.
; APPLICANT: Mooney, Brian P.
; TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND
; TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO
; TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANOATE BIOSYNTHESIS IN PLANTS
; FILE REFERENCE: UMO 1482
; CURRENT APPLICATION NUMBER: US/09/108,020A
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 60/051,291
; EARLIER FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 60/055,255
; EARLIER FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/076,544
; EARLIER FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 54
; SEQ ID NO 15
; LENGTH: 1450
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-108-020-15

Query Match   3.9%;   Score 38.6;   DB 3;   Length 1450;
Best Local Similarity   46.0%;   Pred. No. 0.091;
Matches 168;   Conservative   0;   Mismatches 194;   Indels   3;   Gaps   1;

Qy   632 tgatgtcagtgttgcggtactcctgcaggactcatcacacctattgtgtttaatgc 691
           | ||| | || || | |||||| || | || || |||| | | |||| | ||||
Db   1062 taatattggagttgcaatggccactgaacatggccttgtcgttcctaatataagaatgt 1121

Qy   692 acatataaaggagtggaaaccattgctaatgatgtgttgttttcttagcaaccaaagcaag 751
           | | ||||| || | ||| ||||||| | | ||| |||| ||||| || |||
Db   1122 tcagtcattatctgctagagatataacaaagagctgtcccggttacacacattggcagc 1181

Qy   752 agaggtaaactacagccacatgaattccaggtggtggcactttacgatctccaatttagg 811
           | | || || | | | ||| || || | ||| |||| | | | |||
Db   1182 aaacacaaacttaaccccgaggatgtgactggtggaaccatactcgagtaacattgg 1241

Qy   812 aatgtttggaattaagaattttctctgctattattaaccacctcaagcatgtattttggc 871
           | ||| ||| |||| || |||| | || || | ||| || |||| ||| | |
Db   1242 agcaattggtggtaaattcggatccctctctttaaacttaccggaagttgcaatcatcgt 1301 .

Qy   872 aattggtgcttcagagggaataaaactggtccctgcgataatgaaaaagggtttgatgtggc 931
           | || | || ||| | | |||| || | || || | || | ||| | | ||||| |||
Db   1302 tcttggaagaatcgagaaagttccaaaattctcaaaaagaaggaactgtctatcctgcatc 1361

Qy   932 tagcatgatgtctgttacactcagttgtgatcaccggtggtggatggagcagttggagc 991
           | | ||| || |||| | |||| | | | ||||| | || | | | || || |
Db   1362 gataatgatg--gttaacattgctgcggatcatagagttctagatgggggcaacggtagc 1418

Qy   992 ccagt 996
           | ||
Db   1419 tcggt 1423

Search completed: August 13, 2002, 19:31:35
Job time: 6827 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: August 13, 2002, 17:33:53 ; Search time 1695.31 Seconds
(without alignments)
7969.311 Million cell updates/sec

Title:
Perfect score: 1001
Sequence: 1 atatataagcattgctgac..........cagttggagccagtggctt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

.al number of hits satisfying chosen parameters: 27472414

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :
```
          EST:*
    1:  em_estba:*
    2:  em_esthum:*
    3:  em_estin:*
    4:  em_estmu:*
    5:  em_estov:*
    6:  em_estpl:*
    7:  em_estro:*
    8:  em_htc:*
    9:  gb_est1:*
   10:  gb_est2:*
   11:  gb_htc:*
   12:  gb_gss:*
   13:  em_gss_hum:*
   14:  em_gss_inv:*
   15:  em_gss_pln:*
   16:  em_gss_vrt:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result           %    Query
No.   Score  Match  Length  DB  ID         Description
----  -----  -----  ------  --  ---------  --------------------------------
c  1   739   73.8   751      9  AL040276   AL040276 DKFZp434K
c  2   712.2 71.1   745      9  AL581280   AL581280 AL581280
   3   685.4 68.5   738      9  AU126868   AU126868 AU126868
c  4   682.2 68.2   833      9  AU127045   AU127045 AU127045
   5   630.6 63.0   673     10  BG778185   BG778185 60266342
   6   630   62.9   833     10  BI092658   BI092658 602858446
   7   609.6 60.9   901     10  BI730459   BI730459 603350793
   8   595   59.4  1030     10  BM463514   BM463514 AGENCOURT
   9   582.4 58.2   857     10  BG253201   BG253201 60236415B
c 10   555.6 55.5   720     10  BF939962   BF939962 nac65a04.
  11   552.8 55.2   680     10  BG503419   BG503419 60250846
  12   532.4 53.2  1125      9  AA867881   AA867881 vx22c10.r
  13   526   52.5   664     10  BG390485   BG390485 60241613G
  14   515.4 51.5   654     10  BE850747   BE850747 ux01e11.y
  15   509   50.8   774     10  BI658188   BI658188 60328529G
  16   508   50.7   669     10  BE627518   BE627518 uu52f03.y
  17   473.2 47.3   828     10  BF209700   BF209700 60187423G
```

```
  18   451   45.1   790     10  BI836946   BI836946 60308426
  19   441.8 44.1   574     10  BG087041   BG087041 H31330G11-
  20   433.4 43.3   471     10  N99373     N99373
  21   428.4 42.8   708      9  AW950165   AW950165 EST362130
  22   414.4 41.4   444      9  AU160458   AU160458 AU160458
  23   407.6 40.7   909     10  BE784101   BE784101 60147130G
  24   405.2 40.5   911     10  BF796701   BF796701 60225832G
  25   403.6 40.3   601     10  AA445205   AA445205 v63d04.r
  26   399.4 39.9  1056     10  BM455483   BM455483 AGENCOURT
  27   399.6 39.9   583     10  BF562783   BF562783 U1-R-BJ0p
  28   395   39.5   584     10  AI419467   AI419467 tf16e01.x
c 29   394.2 39.4   575      9  AV667649   AV667649 AV667649
  30   391   39.1   918     10  BF144153   BF144153 601788648
  31   390.4 39.0   563      9  AA275794   AA275794 vc25h09.r
  32   388.4 38.8   582      9  AJ393037   AJ393037 AJ393037
  33   384   38.4   488     10  BE137467   BE137467 ug63g08.y
  34   383.4 38.3   809     10  BI733158   BI733158 603354594
  35   378.8 37.8   463     10  BE629397   BE629397 uu37e01.y
  36   378.6 37.8   567     10  BM236071   BM236071 K0429G04-
c 37   362.4 36.2   463     10  N77305     N77305 yv43e08.r1
  38   361.2 36.1   627     10  BI333861   BI333861 60299938G
  39   360.6 36.0   479      9  AA512816   AA512816 vg13e02.r
  40   360.2 36.0   519      9  AA288380   AA288380 mf47b11.r
  41   358.4 35.8   520      9  AA288386   AA288386 mf47d11.r
  42   355.6 35.5   580     10  AA543438   AA543438 vx37c12.r
  43   355   35.5   564     10  BF725774   BF725774 bx19g09.y
  44   351   35.1   688      9  AL657557   AL657557 AL657557
  45   348.4 34.8   352      9  AA315214   AA315214 EST187073
```

ALIGNMENTS

RESULT   1
LOCUS   AL040276   751 bp   mRNA   linear   EST 29-FEB-2000
DEFINITION AL040276 DKFZp434K2113_r1 434 (synonym: htes3) Homo sapiens cDNA clone
           DKFZp434K2113 5', mRNA sequence.
ACCESSION AL040276
VERSION   AL040276.1   GI:5409233
KEYWORDS   EST.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 751)
  AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE   EST (Koehrer, et al.)
  JOURNAL Unpublished (1999)
COMMENT   Contact: Koehrer K
           MIPS
           Am Klopferspitz 18a D-82152 Martinsried, Germany
           This is the 5' sequence of the clone insert
           Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
           Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
           sequenced by BMFZ (Biomedical Research Center at the Charite,
           Berlin/Germany) within the cDNA sequencing consortium of the German
           Genome Project.
           No s1 sequence available.
           This clone (DKFZp434K2113) is available at the RZPD in Berlin.
           Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
           Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
     source   1..751
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434K2113"
                     /clone_lib="434 (synonym: htes3)"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT   226 a   147 c   160 g   218 t

ORIGIN

Query Match   73.8%;   Score 739; DB 9; Length 751;
Best Local Similarity   99.8%;   Pred. No. 4.6e-181;
Matches   750; Conservative   0; Mismatches   0; Indels   1; Gaps   1;

```
Qy   216 gatcttacacaagtaaaagggacaggacccagatggtagaatcaccaagaaggaatatcgac 275
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GATCTTACACAAGTAAAAGGGACAGGACCAGATGGTAGAATCACCAAGAGGAATATCGAC 60

Qy   276 tcctttgtgcctagtaaagttgtcctcgtccgcgcagctgtcttgcctcccacaggtcct 335
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCTTTTGTGTCCTAGTAAAGTTGTCCTCGTCCGCGCAGCTGTCTTGCCTCCCACAGGTCCT 120

Qy   336 ggaatggcaccagttcctcacaggtgtcttcacagatatccaatcagcaacattcgtcgg 395
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGAATGGCACCAGTTCCTCACAGGTGTCTTCACAGATATCCCAATCAGCAACATTCGTCGG 180

Qy   396 gttattgcacagcgattaatgcaatcaaagcaaaccaatcctacctcattattaccttctatc 455
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTTATTGCACAGCGATTAATGCAATCAAAGCAAAACCATACCTCATTATTACCTTCTATC 240

Qy   456 gatgtaaatatgggagaagttttgttggttacgggaagaacttaataagatattagaaggg 515
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATGTAAATATGGGAGAAGTTTTGTTGGTTACGGGAAGAACTTAATAAGATATTACCTTCTATC 300

Qy   516 agaagcaaaattctgtcaatgactcatcataaaagctcaagctttggcatgttttaaaa 575
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGAAGCAAAATTCTGTCAATGACTCATCATAAAAAGCTTCAGCTTTGGCATGTTTTAAAA 360

Qy   576 gttcccgaagcaaatctcttcttcctgatggacacagttatagacaaaatcatgttgttgat 635
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTTCCCGAAGCAAATCTCTTCTTCGTTGGAGGACACAGTTATATAAGACAAAATCAGTTGTTGAT 420

Qy   636 gtcagtgttcggtcagtactccctgcaggaactcacacagttcacacctatgtgttaatgcacat 695
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTCAGTGTTCGGTCAGTACTCCCTGCAGGACTCATCACAGTTCATCACCTATGTGTTTAATGCACAT 480

Qy   696 ataaaaaggagtggaaaccattgctaatgatgttgttcttagcaacaagaacgaagagag 755
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATAAAAAGGAGTGGAAAACCATTGCTAATGATGTTGTTCTTTAGCAACAAGAACGAAGAGAG 540

Qy   756 ggtaaactacagccaccatgaattccaggtggcacttttacgatctccaatttaggaatg 815
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GGTAAACTACAGCCACCATGAATTCCAGGGTGGCACTTTT-CGATCTCCAAATTAGGAATG 599

Qy   816 tttggaattaagagaatttctctgctattattaaccaccctcaagcatgtatttttggcaatt 875
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   600 TTTTGGAATTAAGAGAATTTCTCTGCTATTATTAACCACCTCAAGCATGTATTTTTGGCAATT 659

Qy   876 ggtgcttcagagaggataaaactggtccctgcaggataatgaaaaagggtttgatgtggctagc 935
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   660 GGTGCTTCAGAGGATAAAACTGGTCCCTGCAGATAATGAAAAAGGGTTTGATGTGGCTAGC 719

Qy   936 atgatgtctgttacactcagttgtgatcacc 966
         ||||||||||||||||||||||||||||||||
Db   720 ATGATGTCTGTTACACTCAGTTGTGATCACC 750
```

Query Match   71.1%;   Score 712.2; DB 9; Length 745;
Best Local Similarity   98.5%;   Pred. No. 4.2e-174;
Matches   708; Conservative   9; Mismatches   2; Indels   0; Gaps   0;

```
Qy   283 tgcctagtaaagttgtcctcgtccgcgcagctgtcttgcctcccacaggtcctggaatgg 342
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   745 TGCCTAGTAAAGTTGTCCTCGTCCGCGCAGCTGTCTTGCCTCCCACAGGTCCTGGAATGG 686

Qy   343 caccagttcctcacaggtgtcttcacagatatccaatcagcaacattcgtcgggttattg 402
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   685 CACCAGTTCCTCACAGGTGTCTTMACAGATATCCCAATCAGCAACATTCGTCGGGTTATTG 626

Qy   403 cacagcgattaatgcaatcaaagcaaaccaatcctacctcattattacctttctatcgatgtaa 462
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   625 CACAGCGATTAATGCAATCAAAGCAAAACCATACCTCATTATTACCTTTCTATCGATGTAA 566

Qy   463 atatgggagaagttttgttggttacgggaagaacttaataagatattagaagggagaagca 522
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   565 ATATGGGAGAAGTTTTGTTGGTTACGGGAAGAACACTTAATAMGRTATTCGAAGGGAGAAGCA 506

Qy   523 aaatttctgtcaatgactcatcataaaagctcaagctttggcatgttttaaaaagttcccg 582
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   505 AAATTYCTYGTCAATGACTTCATCATAAAAAGCTTCAGCTTTGGCATGTTTTAAAAGTTCCCG 446

Qy   583 aagcaaaatttcttcttcttgatggacacagttatataagacaaaatcatgttgttgatcagtg 642
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   445 AAGCAAAATTTCTTCTTCTTGGATGGACACAGTTATAAGACAAAAATCATGTTGTTGATGTCAGTG 386

Qy   703 gagtggaaaccattgctaatgatgttgttctttagcaacaagaacgaagaggtaaac 762
         |||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   385 TYGCGGTCAGTACTCCCTGCAGGACTCATCACACCTATTTGTTGTTAAATGCACATATAAAAG 326

Qy   763 tacagccaacatgaattccaggtgtggcacttttacgatctccaatttaggcaatggaatggaa 822
         |||||||||||||||||||||||||||||:||:|:||||||||||||||||||||||||
Db   325 GAGTGGAAACCATTGCTAATGACTTCAATGATGTTGTTTCTTTMGCAACCAAAGCAAGAGAGAGGTRAAC 266

Qy   823 ttaagaatttctctgctattaaccacctcaagcatgtatttggcaattggtgctt 882
         ||||||||||||||||||||||||||||:|||||||||||||:|||||||||||||||||||
Db   205 TTRAAGAAATTTCTCTGCTATTATTAACCACCTCAAGCATGTATTTTGGCAAATTGGTGCTT 146
```

```
Qy    883  cagaggataaactggtccctgcagataatgaaaaaggggttttgatggggctaagcatgatgt  942
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145  CAGAGGATAAACTGGTCCCTGCAGATAATGAAAAAGGGGTTTGATGTGGCTAGCA1GATGT   86

Qy    943  ctgttacactcagttgtatcaccgggtggtggtggaggcagttcccagtggtggctt      1001
           ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db     85  CTGTTACACTCAGTTGTGATCMCCGGGTGGTGGAGGACGAGTTGGAGCCCCAGTGGCTT   27

RESULT  3
LOCUS       AU126868                                738 bp    mRNA    linear    EST  23-OCT-2000
DEFINITION  AU126868 NT2RP2 Homo sapiens cDNA clone NT2RP2000113 5', mRNA
            sequence.
ACCESSION   AU126868
VERSION     AU126868.1    GI:10951584
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..738
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RP2000113"
                     /clone_lib="NT2RP2"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                     cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT        226 a        171 c        141 g        192 t        8 others
ORIGIN

Qy     21  tataggccaacgaagtaacagattaaaacacacagtgccacacactacccacccccg       80
           |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db      1  TATAGCCAACCGAGCTAACACAGATTTAAAACCACACAGCCACCTACCCCACCCCCG     60

Qy     81  gtggccgctgttcctccaactcccagacttttagtctcctacaacttcagcaacctgccca  140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GTGGCGCCGGTTGTTCCTCCAACTCCCCCAGCCTTTAAAACCACAACCTTCAGCACCTGCCCA  120

Qy    141  gctactccctgctggaccaaaggagggaagggtgtgttttgttgatgcaaaagaaagttggca  200
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  GCTACTCCTGCTGGACCAAAAGGAGGGAAGGGTGTTGTTTTTTAGCCCTTGTCAAAGAAAGTTGGCA  180

Qy    201  gtagagaaaggggattgatcttacacacaagtaaaagttgtagaataccc             260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GTAGACAAAAGGATTCATCTTACACAAGTAAAAGGACAGGACAGATTGGTAGAAATCACC  240

Qy    261  aagaaggatacgactctttgtgctagtaagttgctcctgctccggcgcagctgttgtg    320
```

```
Db    241  AAGAAGGATATCGACTCTTTTGTGCCTAGTAAAGTGGTTTTGTTTTCCTGGTCCTCCGGCTCCGGCAGCTGTTGTG  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy    321  cctcccacaggtcctggaatggcaccagttcctcacaggtgtcttcacaggatatcccaatc  380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CCTCCCACAGGTCCTGGAAATGGCACCAGTTCCTCACAGGTGTCTTCACAGATATCCCAATC  360

Qy    381  agcaacattcgtcggttattgccacagcgattaatgcaaatcaaagcaaacctacctcat  440
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  AGCAACAATTCGTCGGGTTATTGCCACAGCGATTAATCAAAAGCAAACCTACCTCAT  420

Qy    441  tattacctttctatcgaatgtaaaatagggagagaagtgttttgtcggtacggggaaagaaacttaat  500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  TATTACCTTTCTATCAAATGTAAAATATGGGANAANTTTTGTTIGGTTACGGGAAAGAAACTTAAT  480

Qy    501  aagatattaagaagggagagaagcaaaatttctgtcaatgactcaatcaatcaaaagcttcagct  560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  AAGATATTAAGAAGGGANAAGCAAAAATTTCTGTCAAATGACTTCATCAATAAAAAGCTTCAGCT  540

Qy    561  ttggcatgttaaaagttcccgaagcaaattcttcttggatggacacagttataagacaa  620
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    541  TTGGCCATGTTAAAAGTTCCCGAAACAAAATTCTTCTTCGTTGGATGGACACAGTTATGAGAACAA  600

Qy    621  aatcatgttgttgatgtcagttgcgggtcagtactccctgcaggaccatcatcaacacctatt  680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  AATCATGTTGTTGATGTCAGTTGCGGGTCAGTACTCCCTGCAGNAATCATCACACCTATT  660

Qy    681  gtgtttaatgcacat--ataaaaggagtggaaaccattgctaatg-atgttgttcctta  737
           |||||||||||||||   |||||  |||||||||||||||| ||||||||||||||||||
Db    661  GTGTTTAAATGCACATTATATAAAAGGAGAATTGGAAAACCATTGCTAAATGAATNTNTTTCTTTA  720

Qy    738  gcaacacaaaagcaagaga  754
           |||||||||||||||||||
Db    721  NCAACCAAAANCAAGAAA  737
```

```
RESULT  4
LOCUS       AU127045                                833 bp    mRNA    linear    EST  23-OCT-2000
DEFINITION  AU127045 NT2RP2 Homo sapiens cDNA clone NT2RP2000456 5', mRNA
            sequence.
ACCESSION   AU127045
VERSION     AU127045.1    GI:10951761
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 833)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..833
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RP2000456"
                     /clone_lib="NT2RP2"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
```

```
Query Match    68.5%; Score 685.4; DB 9; Length 738;
Best Local Similarity  97.0%; Pred. No. 3.8e-167;
Matches   715; Conservative  0; Mismatches  19; Indels  3; Gaps  2;
```

RESULT 5

```
LOCUS       BG778185                 673 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602666342F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4806083 5',
            mRNA sequence.
ACCESSION   BG778185
VERSION     BG778185.1  GI:14048502
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 673)
  AUTHORS   National Institutes of Health, Mammalian Gene collection (MGC)
  TITLE     Unpublished (1999)
  JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation by: The I.M.A.G.E. Consortium, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1655  row: c  column: 12
            High quality sequence stop: 667.
FEATURES             Location/Qualifiers
     source          1..673
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4806083"
                     /clone_lib="NIH_MGC_60"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccatatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T).  Average
                     insert size 1.5 kb (range 0.9-4.0 kb), 14/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
```

```
BASE COUNT    193 a   168 c   139 g   173 t
ORIGIN
```

Query Match   63.0%; Score 630.6; DB 10; Length 673;
Best Local Similarity   99.0%; Pred. No. 6.2e-153;
Matches 666; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

```
Qy   28 caaccgtagtaacagattaaaaccacacaagtgccaccacctaccccacccccgtggccg  87
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CAACCGGAGTAACAGATTAAAACCACACAAGGCGCCACCACCTACCCACCCCCGGTGGCCG  60

Qy   88 ctgttccccaactcccacccagccttagctcctacacttcagcacctgcccagccagctactc 147
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTGTTCCCCAACTCCCACCCAGCCTTAGCTCCTACACTTCAGCACCCTGCCCAGCCAGCTACTC 120

Qy  148 ctgctggaccaaagggaagggtgtgtttgttagccctcctgcaaagaagttggcagtagaga 207
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CTGCTGGACCAAAGGGGAAGGGGTGTGTTTGTTAGCCCTCCTGCAAAGAAGTTGGCAGTAGAGA 180

Qy  208 aagggattgatctctacacaagtaaaagggacaggaccagatggtagaatcaccaagaagg 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AAGGGATTGATCTTTACACAAGTAAAAGGGACAGGACCAGATGGTAGAATCACCAAGAAGG 240

Qy  268 atatcgactccttttgtgcctagtaaagttgcctccgcccgcagcgtgttgttgcctccca 327
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ATATCGACTCTTTTGTGCCTAGTAAAGTTGCCTCCGCCCGCAGCGTGTTGTTGCCTCCCA 300

Qy  328 caggtcctggaatggcacacagttcctacaggtgtcttcacagatatcccaatcagcacaca 387
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CAGGTCCTGGAATGGCACACAGTTCCTACAGGTGTCTTCACAGATATCCCAATCAGCACACA 360

Qy  388 ttcgtcgggttattgcacagcgttaatgcaataaaacatccattattacc 447
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTCGTCGGGTTATTGCACACGGTTAATGCAATCAAAACCATACCTCATTATTACC 420
```

Query Match   68.2%; Score 682.2; DB 9; Length 833;
Best Local Similarity   95.7%; Pred. No. 2.7e-166;
Matches 711; Conservative 0; Mismatches 30; Indels 2; Gaps 1;

```
BASE COUNT    249 a   192 c   165 g   220 t   7 others
ORIGIN
```

/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

```
Qy   21 tataggccaaccgaagtaacagattaaaaccacacaagtgccaccacctaccccacccccg  80
        |||||||||||||| |||||||||||||||||||||||| | ||||||||| ||||||||
Db    1 TATAGGCCAACCGAGTAACAGATTAAAACCACACAAGGCGCCACCACCTACCCACCCCCG  60

Qy   81 gtggccgctgttccccaactcccacccagccttagctcctacacttcagcaccctgccca 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   61 GTGGCCGCTGTTCCCCAACTCCCACCCAGCCTTAGCTCCTACACTTCAGCACCCTGCCCA 120

Qy  141 gctactcctgctggaccaaagggggaagggtgtgtttgttagccctcctgcaaggaagttggca 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GCTACTCCTGCTGGACCAAAGGGGAAGGGGTGTGTTTGTTAGCCCTCCTGCAAAGAAGTTGGCA 180

Qy  201 gtagagaaagggattgatctctacacagtaaaagggacaggaccagatggtagaatcacc 260
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTAGAGAAAGGGATTGATCTTTACACAAGTAAAAGGGACAGGACCAGATGGTAGAATCACC 240

Qy  241 aagaaggataatcgactctttgtgcctagttaaaagtgctcaatgactttgtgttgttgg 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AAGAAGGATATCGACTCTTTTGTGCCTAGTAAAGTTGCCTCCGCCCGCAGCGTGTTGTTG 300

Qy  261 aagagaaggatatcgactcttttgtgcctagtaaaggacaggaccagatggtagaatcacc 320
Db  181
```

RESULT 6

```
Qy    448 tttctatcgatgtaatatgggaggagaagttgtgttggtacggaaaagaacttaatagaat 507
          ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    421 TTTCTATrCGATrGTACrATATGGGAGAAG-TTTGTTGTTGGTACGGAAAGAACTTAATAGAATAT 479

Qy    508 tagaagggagagaagcaaaatttcctgtcaatgacttcatccataaaagcttcagcttt 567
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    480 TAGAAGGGAGAAGCAAAGCAAAAATTTCTGTCAATGACTTCATCATCAT AAAAGCTTCAGCTTTGGCAT 539

Qy    568 gtttaaaagttcccgaagcaaaatcttcttggatggaacacagtgttataagacaa-aatcat 626
          ||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    540 GTTTAAAAGTTCCCGAAGCAAAATTCTTCTTGGATGGAACACAGTGTTATAAGACAACAATCAT 599

Qy    600 GTTGTCCGACrGTGTCAGTGTTGCGGTCAGTACTCCTCTGCAGGACTCATCACACCTATTGTGTT 659
          || || || |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    627 gttgttga-tgtcagtgtgcgggtcagtactccctgcaggactccatcacacctattgtgtt 685

Db    592 cttccttggatggaacacagtgtt ataagacaacaaaatcatcat gttgtcc 651
Db    540 GTTTAAAAGTTCCCGAAGCAAAATTCTTCTTGGATGGAACACAGTGTTCCCGGAAGCAAAATT 240
Db    600 GTTGTCCGACrGTGTCAGTGTTGCGGTCAGTACTCCTCTGCAGGACTCATCACACCTATTGTGTT 659

Qy    686 taatgcacatata 698
          ||||| |||| ||
Db    660 TAATGCACATTTA 672
```

Query Match   62.9%; Score 630; DB 10; Length 833;
Best Local Similarity   98.9%; Pred. No. 9.6e-153;
Matches   645; Conservative   0; Mismatches   5; Indels   2; Gaps   1;

```
Qy    352 ctacagtgtgtcttcacagatatcccaatcagcaacatcgtcgggttattgcacagcgat 411
          |||||| |||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      1 CTACAGGTGTCTTCACAGATATCCCAATCACAGCAACATTCGTCGGGTTATTGCACAGCGAT 60

Qy    412 taatgcaatcaaagcaaccataacctcattattacttcatcgatgtaataatgtgggag 471
```

RESULT 7

```
Db     61 TAATGCAATCAAAGCAAAACCATAACCTACCTCATTATTACCTTCATCrGATGTAAATATGGGAG 120
          |||||||||||||| |||||||||||||| |||||||||| |||||| ||||||||||||||

Qy    472 aagtttgttggtacggaaagaacttaatagatattagaagggagaagcaaaatttctg 531
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    121 AAGTTTGTTGGTACGGAAAGAACTTAATAGATATTAGAAGGGAGAAGCAAAAATTTCTG 180

Qy    532 tcaatgacttcatcataaaagcttcagcttcagctttggcatgttcccgaagcaaatt 591
          ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db    181 TCAATGACTTCATCATAAAAGCTTCAGCTTTGGCATGTTCCCGGAAGCAAAATT 240

Qy    592 cttcttggatggaacacagtgttataagacaacaaaatcatcgttgttcagtgtcca 651
          ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    241 CTTCTTGGATGGAACACAGTTATAAGACAAACAAAAATCATGTTGTTCAGTGTTGTCGGTCA 300

Qy    652 gtactcctgcaggactcatcacacctattgtgttataagtgcacatatataaagagtggaa 711
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||| |||||
Db    301 GTACTCCTGCAGGACTCATCACACCTATTGTGTTTAAGTGCACATATAAAAGGAGGTGGAAA 360

Qy    712 ccattgtaatgatgttgttctttagtcaaccaacaaagcaagagagaggggtaaactacaagccac 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCATTGCTrAATrGATrGATrGTTGTTrCTTTAGTCAACCAAAGCAAGAGAGAGGGGTAAACTACAGCCAC 420

Qy    772 ataaattccaggtgtggcacttttacgtatctccaatttaggaaatgtttggaattaaggaatt 831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGAAAATTCCAGGGTrGGCACTTTTrACGGATCTCCAATTrACGGAAAATrGTTTTGGAATTAAGAAATT 480

Qy    832 tctctgtctattattaacccaacctcaagcatgtcatgtattttggcaattggtgcttcagagggta 891
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTCTGCTrATTATTrAACCCACCTCAAGCATGTCATGTATTTTGGCAATGGGTGTCTTCAGAGGGArA 540

Qy    892 aactggtccctgcagata--atgaaaaaggggtttgatggtggctaagcatgtgatgtctgttac 949
          ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db    541 AACTGGTCCCTGCAGATATGGAACACACAGGGTTCGATGTGGGCTrAGCATGATrGTCTGTTrAC 600

Qy    950 actcagttgtgatcaccggggtggttgggtggggagcaggttggagccccagttggcttt 1001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ACTCAGTTGTGATrCACCGGGTGGTTGGGArAGGAGCAGAGTTGGGAGCCCAGTGGGCTT 652
```

Left column

```
ORIGIN
BASE COUNT    243 a    196 c    225 g    237 t
                    Note: this is a NIH_MGC Library."
                    full-length clones and constructed by Life Technologies.
                    Average insert size 3.3 kb. Library enriched for
                    Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                    /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                    /lab_host="DH10B (phage-resistant)"
                    /tissue_type="retina"
```

Query Match
Best Local Similarity   60.9%;
                         86.4%;   Pred. No. 2e-147;   DB 10;   Length 901;
Matches 731; Conservative   0;   Mismatches 109;   Indels   6;   Gaps   5;

```
Qy  162 ggaagggtgttgttagcccctcctttggcagtagagaagagaaagggattgatctt    221
        |||||| |||  | |||| ||| ||||  | ||||| ||| ||| |||||||| ||
Db    1 GGAAGGGTGGTTCGTTAGTCCTCTTGCCAAGAAGTTGGCAGCAGAGAAAAGGGATTGACCTC   60

Qy  222 acacagtaaaagggacaggagacaagtggtaggaatcaccagagaagagaatatcgactcttt   281
        |||| ||| |||  || |||  |  || ||||  | ||||| ||||| || |  || | ||||
Db   61 ACCAAGTTAAAGGGACCAGGAGACCAGAATCATCAAGAAGGACATTGACTCTTTT        120

Qy  282 gtgcctagtaaagttgctcctgctccggcagcgtgttgtgcctcccacagtcctggaatg   341
        ||||| ||| ||| ||| ||  || |||||||| ||| | || || ||| || ||||| |
Db  121 GTGCCTTCTAGGCTGTCTGCCTCGAGCACGTGCCATGGCTCTCCCGGGTCCCAAGAGTG    180

Qy  342 gcaccagttcctacaggtgtcttcacagatatcccaatcagcgaacattcgtcgggttatt    401
        |||| ||||||||||||  | ||| ||| |||||||| ||||||| ||||| |  |  |||
Db  181 GCACCAGCTCCTGCAGGTGTCTTCACAGATCCCCATCAGCGAACATTCGTCGGGTTATT     240

Qy  402 gcacacgcgattaatgcaaatcaaaccatacaccctcattcttctatctcgatgta    461
        ||||| || || ||||||||| || || ||| | |||  |||| ||| |  |  ||
Db  241 GCGCAGAGGCTCATGCAAGCAGAGACCATACCTCCTCATTTCATCAGATGTA        300

Qy  462 aatatgggagaagtttgttcggtacggaaagagaacttaataagatattagaaggagaagc   521
        || |||||| ||| |||| |  | ||| || ||||| ||  | |||| | |||| ||||||
Db  301 AATATGGGAGAGGTGTGTTCGGTACGGAAGGAGAACTTAATAGATATTAGAAGGGAGAAGC    360

Qy  522 aaaaatttctgtcaatgactccatcataaaagcttcagctttggcatgtttaaagttccc   581
        ||||| ||||  ||| |||  || | |||||| || ||  ||||| ||  ||||| ||||
Db  361 AAAAATTCCGTCAACGACTTCATTATAAAGGCTTCAGCTTTGGCCTGTCTCTGAAAGTTCCC   420

Qy  582 gaagcaaattcttcttgggattggaacaacagttataagacaaaatcaattgttgatgtcagt   641
        |||||||| ||||||||  | |||  || | |||| ||| ||| || |  | |||| |||||
Db  421 GAAGCAAACTCGTCTTGGATGGAACAGCAGTTATAAGGCAAAAATCATGTTGGTTGACGTCAGT    480

Qy  642 gttgcggtcagtactcctgcaggactcatcacaccctattgtgttaatgcacatat-aaa    700
        |||| |||||| | |||| ||| ||| || | |||| ||||| ||| | ||||||| |||
Db  481 GTTGCTGTCAGTGACTACCCCTGCAGGACTTATCACCCCTATTGTGTTTAATGCACATAAAA    540

Qy  701 aggagtggaaacccattgctaatgatgttgttcttcagcaaccaaagcaagagaggtaa   760
        ||||| ||||| ||||| |  ||||| ||||| |||| |||||||||||| |||| ||
Db  541 AGGACTGGAAACCATTCCTAGGATGTTGTTCTTCTTTAGCGAGAGGAGGGGTAA        600

Qy  761 actacagccacaatgaattccagggggtcg--cactttacgaatctccaattaggaatgtttg   819
        |||| ||| ||| |||||  || || || |  |||||| |  |||  ||||||| | ||||
Db  601 ACTTCAGCCTCATGAGTTCCAGGGTCGGAAACATTTACAAATCTCCAACTTAGGGATGTTCG    660

Qy  820 gaattaagaaatttcctgctattattaacccacctcaagcatgtatttggcaattggtg   879
        |||| || ||| |||| |  ||| || || | |||||||||| | |||||| |||
Db  661 GGGAGAAGTTTGTTGCTATTATTAACCCACCTCAAGCATGTATTTGGCAATGGTG        720

Qy  880 cttcagaggagatgaaactggtccctgcaga--taatgaacagagtggctagcatg   938
        || |||||| || ||||| || |||||||||  |  |||||| || ||| ||  ||
Db  721 CTTCAGAGGATGAAACTCATCATCAATAAAAGCTTCAGCTTTGATGTGGCTAGTGTG    780

Qy  939 atgt-ctgttacactcagttgatcaacg--ggtcggtggatggacagttgggagcccag   995
        |||| ||||||| |||||| |||| | |  |||| ||||||||| |||||||||||||||
Db  781 ATGTCCCGTTACACTCAGCGTGTGATCATCGAAGTGTTGGATGGACGGTTGGAGCCCAG     840

Qy  996 tggctt  1001
        ||||||
Db           840
```

Right column

```
Db  841 TGGCTT  846
```

RESULT   8

```
LOCUS       BM463514    1030 bp    mRNA    linear    EST 05-FEB-2002
DEFINITION  AGENCOURT_6433730 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5585926
            5', mRNA sequence.
ACCESSION   BM463514
VERSION     BM463514.1   GI:18512556
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1030)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12352  row: p  column: 23
            High quality sequence start: 17
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1..1030
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5585926"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT    302 a    205 c    209 g    314 t
ORIGIN
```

Query Match
Best Local Similarity   59.4%;
                        100.0%;   Pred. No. 1.3e-143;   DB 10;   Length 1030;
Matches 595; Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy  407 gcgattaatgcaatcaaagcaaacctacctcattactaccttctatctgatgtaaatat   466
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 GCGATTAATGCAATCAAAGCAAACCTACCTCATTACTACCTTCTATCTGATGTAAATAT    90

Qy  467 gggagaagtttgttggtacggaaagagaactaataagatattagaaggagaagcaaaat   526
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91 GGGAGAAGTTTGTTGGTACGGAAAGAGAACTTAATAAGATATTAGAAGGAGAAGCAAAAT   150

Qy  527 ttctgtcaatgactccatcataaaagcttcagctttggcatgtttaaagttcccgaagc   586
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  151 TTCTGTCAATGACTCCATCATAAAAGCTTCAGCTTTGGCATGTTTAAAGTTCCCGAAGC   210

Qy  587 aaattcttcttgggatcacagttataagacaaaatcaatcatgtttgatgtcagtgttgc   646
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  211 AAATTCTTCTTGGGATCACAGTTATAAGACAAAATCAATCATGTTTGATGTCAGTGTTGC   270

Qy  647 ggtcagtactcctgcaggactcatcacaccctattgtgttaatgcacatataaaaggagt   706
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  271 GGTCAGTACTCCTGCAGGACTCATCACACCCTATTGTGTTAATGCACATATAAAAGGAGT   330

Qy  707 ggaaaccattgctaatgatgttgttcttcagcaaccaaagcaagagagggtaaactaca   766
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      766
```

```
Db   331 GGAAACCAATTCCTAATGACTGGTTGTTGTTCTTTAGCACAACCAAAGACAAGAGAGGGTAAACTACA 390
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   767 gccacatgaattccaagggtggcgcacttttacgatctccaattaggagataatagtttggaattaa 826

Db   391 GCCACATGAATTCCAGGGTGGCGCACTTTTACGATCTCCAATTGAGAGTTGGAATTAA 450
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   827 gaattcctcgctacttttattaaccccacccaagacatgtattttggcaattgagagttggcttcaga 886

Db   451 GAATTTCTCTGCTATTATTAACCACCACCCAAGCATGTATTTTGGCAATTGGTGCTTCAGA 510
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   887 ggataaactggtcccctgcagataatgaaaaagggtttgatgtggctagacatgattctgt 946

Db   511 GGATAAAACTGGTCCCCTGCCAGATAATAATGAAAAAAGGGTTTGATGTGGCTAGCCATGATGTCTGT 570
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   947 tacactcagttgtgatcaccgggtggtggatggagcccagtggtt 1001

Db   571 TACACTCAGTTGTCTGATCACCCGGGTGGTGGGGATGGAGCAGTGGGCCCAGTGTGGCTT 625
```

```
JULT    9

LOCUS      BG253201      857 bp   mRNA    linear   EST 13-FEB-2001
DEFINITION 602364158P1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4472391 5',
           mRNA sequence.
ACCESSION  BG253201
VERSION    BG253201.1  GI:12763017
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 857)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM10293  row: c  column: 16
           High quality sequence stop: 704.
FEATURES            Location/Qualifiers
     source          1..857
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4472391"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT  284 a   185 c   156 g   232 t
ORIGIN
```

```
Query Match     58.2%;  Score 582.4;  DB 10;  Length 857;
Best Local Similarity   99.7%;  Pred. No. 2.2e-140;
Matches 594;  Conservative    0;  Mismatches  1;  Indels  1;  Gaps   1;

Qy   407 gcgattaatgcaatcaaagcaaaccatacctcattattacctcttctatcgataaaatat 466
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GCGATTAAATGCAAATCAAAAGCAAACCATACCTCATTATTATTACCTTTCTATCAATGTAAAAAT 60
Qy   467 gggagaagtttggtggtacggaaggaacttaatagatattagaagggaagcaaaat 526
```

```
RESULT 10

LOCUS      BF939962/c       720 bp    mRNA    linear    EST 22-JAN-2001
DEFINITION nac65a04.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3439014 3'
           similar to SW:ODP2_HUMAN P10515 DIHYDROLIPOAMIDE ACETYLTRANSFERASE
           COMPONENT OF PYRUVATE DEHYDROGENASE COMPLEX, MITOCHONDRIAL
           PRECURSOR ;, mRNA sequence.
ACCESSION  BF939962
VERSION    BF939962.1  GI:12357282
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 720)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute / National Institute of Neurological
           Disorders and Stroke, Brain Tumor Genome Anatomy Project
           (CGAP/BTGAP), Tumor Gene Index
JOURNAL    Unpublished (1998)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
           Ph.D.
           cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
           Bonaldo, Ph.D.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL, send email to:
           info@image.llnl.gov
           Seq primer: -40UP from Gibco
           High quality sequence stop: 464.
FEATURES            Location/Qualifiers
     source          1..720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
Db    61 GGGAGAAGTTTTGTTGGTACGGAAGGAAGAACTTAATAGATATTAGAAGGGAGAAGCAAAAT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    61 tttcgtcaatgactcatcataaaagcttcagctttggcgatgttcatgaaaagttcccgaagc 120

Db   121 TTCTGTCAATGACTTCATCATAAAAAGCTTCAGCTTTGGCATGTTCATGAAAAGTTCCCGAAGC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   527 aaattcctcttggatggacacagttataagacaaaatcatgtgtattgtcagtgttgc 646

Db   181 AAATTCTTCTTGGATGGACAGAGTTATAAGACAAAATCATGTGTTGTTGATGTCAGTGTTGC 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   587 ggtcagtactcctgcaggaactcatcaacacctattgttgtttaatgcaacatataaaggagt 706

Db   241 GGTCAGTACTCCTGCGCAGGAACTCATCATCAACACCTATTGTTGTTTAATGCAATATAAAAGGAGT 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   647 ggaaaccatgttgctaatgatgttgttcttttagcaaaacaaagcaaagagaggggtaaactaca 766

Db   301 GGAAACCAATTGCTAATGATGTTGTTCTTTTAGCAAACAAAGCAAAGAGAGAGGGTAAACTACA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   707 gggaaaccatgttgctaatgatgttgttcttttagcaacaaacaaagcaaagagaggggtaaactaca 826

Db   361 GCCACATGAATTCCAGGGTGGCACTTTTACGATCTCCAATTTACGAATCGTTTAGGAATGTA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   827 gaattcctcgctattattaaccccacccaagacatgtattttggcaattgagagttggcttcaga 886

Db   421 GAATTTCTCTGCTATTATTAACCCACCACCCAAGCATGTATTTGGCAATTGGTGCTTCAGA 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   887 ggataaactggtcctgcagataatgaaaaagggtttgatgtggctagacatgatttctgt 946

Db   481 GGATAAAACTGGTCCCCTGCAGATAATGAAAAAAGGGTTTGATGTGGCTAGCCATGATGATGTCTGT 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   947 tacactcagttgtgatcaccgggtggtggatggagcccagtggtt 1001

Db   541 TACACTCAGTTGTCTGATCACCCGGGTGGTGGGGATGGAGCAGTGGGCCCAGTGTGGCTT 596
```

Left column

RESULT 11

```
LOCUS        BG503419        680 bp   mRNA    linear   EST 27-MAR-2001
DEFINITION   602SS08461 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4658459 5',
             mRNA sequence.
ACCESSION    BG503419
VERSION      BG503419.1  GI:13464936
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
             /clone="IMAGE:3439014"
             /clone_lib="NCI_CGAP_Brn23"
             /tissue_type="glioblastoma (pooled)"
             /lab_host="DH10B"
             /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
             modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
             strand cDNA was primed with a Not I - oligo(dT) primer [5'
             TGTTACCAATCTGAAGTGGGACGCGGCCGCAATATCTTTTTTTTTTTTTTTTTTT
             T3']; double-stranded cDNA was ligated to Eco RI
             adaptors (Pharmacia), digested with Not I and cloned into
             the Not I and Eco RI sites of the modified pT7T3 vector.
             Library is normalized, and was constructed by Bento
             Soares and M.Fatima Bonaldo."
```

```
BASE COUNT      228 a    147 c    123 g    221 t    1 others
ORIGIN
```

```
Query Match             55.5%;  Score 555.6;  DB 10;  Length 720;
  Best Local Similarity 97.8%;  Pred. No. 1.9e-133;
  Matches 564;  Conservative  0;  Mismatches 15;  Indels  0;  Gaps  0;
```

```
Qy  423 aagcaaaccataccccattcattattacctttctatcgatgtaaatatgggagaagaagtttgttg 482
        |||||||| ||||| ||||| |||||||||||||| ||| ||| ||||||||||||||| |||||
Db  720 AAAGCAACAACCATACCCTCATTATAAACTTTCTATCAATGTAATAAATATGGGAGAAGTTTGTTG 661

Qy  483 gtacggaaagaacttaataagatattagaagggagagaagcaaaaattcctgtcaatgactc 542
        ||||| |||| ||| |||||  |||||| |||||||||  |||||||||||||||| ||| |
Db  660 GTACGAAAAGACTTTATANAGATATAGAAAGGGAGAAGCAAAAATTCCTGTCAAATGACTTC 601

Qy  543 atcataaaagcttcagccttggcatgtgttaaaagttccccgaagcaaattcttcttggatg 602
        |||||||||||||| ||| ||||||| |||||||||||| |||||||||||||||||| ||||
Db  600 ATCATAAAAGCTTCACCTTTGGCATGTTTAAAAGTTCCCGAAGCAAATTCTTCTTTGGATG 541

Qy  603 gacacagttataagacaaaatcatgttgttgatgtcagtgttgccggtcagtactcctgca 662
        |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  540 GACACAGTTATAAGACAAAAAATCATGTTGTTGATGTCAGTGTTGCCGGTCAGTACTCCTGCA 481

Qy  663 ggactcatccacccattgtgttaatgcacatataaaaggagtggaaccattgctaat 722
        ||||||||| ||| |||||||| ||||||||||  ||||||||||| |||| ||||
Db  480 GGACTCATCACACCTATTGTGTTTAATGCACATAATAAAAGGAGTGGAAACCATTCCTAAT 421

Qy  723 gatgtgttcctttagcaaccaaagcaaagaggagggtaaactacagccacatgaattccag 782
        ||||| || ||| |||||| |||||||||||| |||||||||||||||||||||||||||||
Db  420 GATGTTGTTCTTCTTTAGCACCAAAGCAAAGAGAGGGTAAACTACAGCCACATGAATTCCAG 361

Qy  783 ggtggcacttttacgatctccaatttaggaattaagaaattctctgtcatt 842
        |||||||||||||| |||||||||||||||| |   |||||||||||| |||
Db  360 GGTGGCACTTTTACGGATCTCCAATTTAGGAATGTTTGGAATTAAGAAATTCTCTGCTATT 301

Qy  843 attaaccacctcaagacatgtatttggcaattggtcgtcttcaagaggagataaactggtccct 902
        ||||||||||||||| ||||||||| |||||||||||  |||||||||| ||||||||||||||
Db  300 ATTAACCACCTCAAGCCATGTATTTTGGCAATTGGTCCTTCAGAGGATAAACTGGTCCCT 241

Qy  903 gcagataataatgaaaaagggtttgatgtggctagcacgtcagttctgttcacctcagtt.tgat 962
        ||||||||| |||||||||||||||||||||||||||| |  ||| ||| |||| | |||| ||||
Db  240 GCAGATAATGAAAAAGGGTTTGATGTGGCTAGCACATGATGTCTGTTACACTCAGTTGTGAT 181

Qy  963 caccggggtggtggatggagcaagtggagcccagtggctt 1001
        ||||||  ||| ||||||||||| ||||||||||||||||
Db  180 CACCGGGTTGGTTGGATGGAGCAGTTGGAGCCCAGTGGCTT 142
```

Right column

```
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 680)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1452 row: 1  column: 12
             High quality sequence stop: 675.
```

```
FEATURES             Location/Qualifiers
     source          1..680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4658459"
                     /clone_lib="NIH_MGC_61"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     Sfil (ggccgcctggcc); Site_2: Sfil (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACGCGTATATGCC-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
```

```
BASE COUNT      201 a    166 c    147 g    166 t
ORIGIN
```

```
Query Match             55.2%;  Score 552.8;  DB 10;  Length 680;
  Best Local Similarity 99.3%;  Pred. No. 9.9e-133;
  Matches 576;  Conservative  0;  Mismatches  2;  Indels  2;  Gaps  2;
```

```
Qy    1 atatatacagcatttgctgactataggccaaccgaagtaacaggaatttaaaaacacaagtgc 60
        ||| | ||||||||||||||||||||||||| ||||||||||||| |||||||||||||| |||
Db  101 ATATGTCAGCATTTGCTGACTATAGGCCCAACCGAAGTAACAGAGATTTAAAAACACAGTGC 160

Qy   61 caccacctacccccccccgtggccgcgtgttcctccaacctccccagcccttagctccta 120
        ||||||||||||| | |||| ||||| |||||||||||| ||||||||||||||||||||
Db  161 CACCACCTACCCCCACCCCCCGGTGGCCCGGTGTTCCTCCAACTCCCCAGCCCTTAGCTCCTA 220

Qy  121 caccttcaagcaccctgcccagctactcctgctggaccaaaggaaagggtgttgttgtagcc 180
        |||||||  |||||||||||||||||||||||||||||||||   |||||||||||||||||
Db  221 CACCTTCACCACCCTGCCCAGCTACTCCTGCTGGACCAAAGGGGCAAGGGTGTTGTTGTAGCC 280

Qy  181 ctcttgcaaagaagttggcagtagaagaaagggattgatcttacacaagtaaagaaggacag 240
        |||||||||||||||||||||||||| ||||||||||||||||||||||||  ||||||||
Db  281 CTCTTGCAAAGAAGTTGGCAGTAGAGAAAGGGATTGATCTTACACAAGTAAAAGGACAG 340

Qy  241 gaccagatggtagaattcaccaagagaggatatcgacactctttgtgcctagtaaagttgctc 300
        ||||||||||||||| |||||||| |||||||||| |||||||||||||||||||||||||||
Db  341 GACCAGATGGTAGAATCACCAGAAGGATATCGAACTCTTTGTGCCTAGTAAAGTTGCTC 400

Qy  301 ctgctccggccagttgttgtctcccacaggtcctggaatggcaacagttctacaggtg 360
        ||||||||||||| ||||||| |||||||||||||||||||||||||||| |||||||
Db  401 CTGCTCCGGCCAGCTGTTGTTCTCCCACAGGTCCTGGAATGGCAACAGTTCCTACAGGTG 460

Qy  361 tcttcacagagatatcccaatcagcaacattcgttcgggttattgcaacagcgattaatgcaat 420
        ||||||||| |||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  461 TCTTCACAGATATCCCAATCAGCAACATTCGTCGGGTTATTGCAACAGCGATTAATGCAAT 520

Qy  421 caaagcaaccatcacctcatta-ttacctttctatcgatgtaaatatgggagaagtttttg 479
        ||||||||||||||||||| || ||||||||||||||||||||||||||||||||||||||
Db  461 CAAAGCAACCATCACCTCATTACTTACCTTTCTATCGATGTAAATATGGGAGAAGTTTTTG 520
```

```
RESULT 12

LOCUS       AA867881                1125 bp    mRNA    linear   EST 16-MAR-1998
DEFINITION  vx22c10.r1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:1265202 5', similar to gb:Y00978_cds1 DIHYDROLIPOAMIDE
            ACETYLTRANSFERASE COMPONENT (HUMAN);, mRNA sequence.
ACCESSION   AA867881
VERSION
  .WORDS    AA867881.1  GI:2963326
            EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1125)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:667754
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1..1125
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1265202"
                     /sex="male"
                     /clone_lib="Soares_thymus_2NbMT"
                     /tissue_type="thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCGAGACTGGAGCGGCCGCGCGTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7D vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       275 a     235 c     305 g     309 t      1 others
ORIGIN

Query Match         53.2%;  Score 532.4;  DB 9;  Length 1125;
Best Local Similarity   86.6%;  Pred. No. 2.4e-127;
Matches 609;  Conservative 0;  Mismatches 92;  Indels 2;  Gaps 2;
```

```
Db   521 CAAACCAAAACCATACCTCATTATTTAACCTTTCTATCGATTGAAATATGGGAGAAGTTCG 
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Qy   480 ttggtacggaaagaacttaataataagatattagaagagggagaagcaaaaattctg    539
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   581 -TGGTACGGAAAGAACTTAATAAGATAATTAGAAGGGGAGAAGCAAAAATTTCTGTCAATGAC   639
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   540 ttcatcataaaagcttcagcgtttggcatgttaaagttc    579
         ||||||||||||||||||||||||||||||||||||||||
Db   640 TTCATCATAAAAGCTTCAGCGTTTGGCATGTTAAAAGTTC   679
```

```
Qy   290 taaagttgctcctgtctccggcagctgtgtgtctcccacagtgtcctggaatggcaccagt 349
         ||| ||  ||||||||||||  ||||| |||||||||||||||||||||||||||||||||
Db     2 TAAGGCTGGTCCTGCCTGCAGCACGTCGCACCTGCCATGGCTCCCCGGGGTCCAAGAGTTTCG  61
Qy   350 tcctacaggtgtcttcacagatatcccaatcagcaaacacaattcgtcgggttcattgcacgcg 409
         |||||||||||||| |||||| ||||||  ||||||| ||||  ||||  ||| ||||||||
Db    62 TCCTGCAGGTGCTTCTCACAGCTATCCCAGACCATCCCATCAGCAACACATCGTCGTGGATTGCGCAG  121
Qy   410 attaatgcaatcaaagcaaaaccataccccatcattactcttctatcatgtaaatatggg 469
         | ||| |||||||||||  |||||| |||| |||||||  ||||||  ||||| ||||||
Db   122 GCTCAGTGCAGTCGAAGCAGGAAGACTATACCTCATTATTACCTTTCTGTCGATTGTAAATATGG 181
Qy   470 agaagtttgttcggtacggagaggaacttaataataagatattagaagggagaagcaaaatttc 529
         ||||| |||| |||||||||||||  |||||||  ||||||  |||||||  ||||||||||
Db   182 AGAGGTGCTGGTTTGGTACGGAAGGAACTTAATAAGATGCTTGAAGGGGAGAAGGAAAATCTC 241
Qy   530 tgtcaatgacttcatcataaaagcttcagcgtttaaaagttcccggaagcaa 589
         ||||||||||| || |||||||||  ||||| ||||| ||||||  |||||
Db   242 CGTCAACGACTTCATTATAAAGGCTTCAGCTTTGGCGTTGAAAGTTCCCGAAGCAAA 301
Qy   590 ttcttctggatggacacagttataagacaaaaatcaatgttgatgtcaagtgtcgt 649
         |||||| ||||||||||||||||||||||  ||||||| ||||||  ||||| ||
Db   302 CTCGTCTTGGATGGACACAGTTATAAGGCAAAAATCATGTGGTTGACGTCAGTGTGTCGT 361
Qy   650 cagtactcctgtcaggactcatcacacatattgtgttaatgcacatataaaggagtgga 709
         |||||| || ||||| |||| ||  |||| ||||| ||||||||| |||| |||||||
Db   362 CAGTACCCCGTGCAGGACTTATCACCCTATTGTGTTAATGCACACATAAAAGGACTGGA 421
Qy   710 aacaattgctaatgatgttgttctttagcaaccaaagcagagaggggtaaactacagcc 769
         |||||| ||||| |||||||||||||| ||||  |||||||| ||||||||||||||||
Db   422 AACACATTGCTAGTGATTGTTGTTTCTTTGACCCTCCAAAGCGGAGAGAGGGTAAACTTCAGCC 481
Qy   770 acatgaattccaggtggcacttttacgtctccaatttaggaatgttttggaattaagaa 829
         ||||| ||| ||||| ||| |||| |||||||||| ||||||| ||| |||| ||||||
Db   482 TCATGAGTTCCCAGCGTGGAACACATTTACAAATCAATCATGGGTGTATGGGAATTTAAGAA 541
Qy   830 ttctctgctattaacccacctcaaggcgatgtatttggcaattggctgcttcagagga 889
         |||| ||  ||||| |||| |||| || |||| ||||||| ||  ||||| |||||||
Db   542 TTTCTCTGCAGATATTAATCCACCTCAGGCCATGTATTTCTCGGCAA-TCGTGGCTTCAGAGGA 600
Qy   890 taaactggtccctgccagataataaagaaaagggttttggatgtggcgacgatgctgttac 949
         ||||  ||| ||||||  ||||| | ||||||| |||| |||  |||||| ||| |||
Db   601 TAAACTGATTCCAGCAGATAATGAGAAAGGCTTTGATGTGGCTAGTGTGAATGTC-GTTAC 659
Qy   950 acctcagtttgtgatcaccgggtgggtcggatgggagcagttggagcc 992
         |||||| |||||||| ||| |||| | |||||| ||||||||||||||
Db   660 ACTCAGCTGTGTGATCATCGAGTTGTCGGATGGAGCAGTTGCGAGCC 702
```

```
RESULT 13

LOCUS       BG390485                664 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602416136F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4524344 5',
            mRNA sequence.
ACCESSION   BG390485
VERSION     BG390485.1  GI:13283933
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
```

Left section

```
FEATURES          found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLAM10428 row: h column: 09
                  High quality sequence stop: 660.
      source    1. .664
                  Location/Qualifiers
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4524344"
                  /clone_lib="NIH_MGC_92"
                  /tissue_type="embryonal carcinoma, cell line"
                  /lab_host="DH10B (phage-resistant)"
                  /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 2.5 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
                      Note: this is a NIH_MGC Library."

BASE COUNT    206 a    113 c    144 g    201 t
ORIGIN
```

```
Qy   453  atcgatgtaatatgggagaagagttttgttggtacggaaagaactaataataagaatattaga 512
          ||||||||||  |||||| ||||||||||||||||||||  |||||| ||||||||||||||||
Db     1  ATCGATGTAAATATGGGAGAAGAGTTTGTTGGTACGGAAAGAACTTAATAAGATATTAGAA    60

Query Match       52.5%;  Score 526;  DB 10;  Length 664;
Best Local Similarity   99.6%;  Pred. No. 9e-126;
Matches  548;  Conservative  0;  Mismatches  0;  Indels  2;  Gaps  2;
```

```
Qy   453  atcgatgtaatatgggagaagagttttgttggtacggaaagaactaataataagaatattaga 512
Db     1  ATCGATGTAAATATGGGAGAAGAGTTTGTTGGTACGGAAAGAACTTAATAAGATATTAGAA    60
Qy   513  gggagaagcaaatttcctgtcaaatgacttcatcataaaaaagcttcagcttgggcatgtta 572
Db    61  GGGAGAAGCAAATTTCTGTCAAATGACTTCATCATAAAAAGCTTCAGCTTGGCATGTTTA   120
Qy   573  aaagttccccgaagcaaattcttcttggatgagtgacacagttataagacaaaatcatgttgtt 632
Db   121  AAAGTTCCCGAAGCAAATTCTTCTTCTGGATGGACACAGTTATAAGACAAAATCATGTTGTT   180
Qy   633  gatgtcagtgttgcggtcagtcactcctgcaggactcatcaccaccattgtgtttaatgca 692
Db   181  GATGTCAGTGTTGCGGTCAGTACTCCTGCAGGACTCATCACCACCTA-TGTGTTTAATGCA   239
Qy   693  catatatataaaaggagtggaaaccacattgctaatgatgatgttgttcttagcaacaacaaga 752
Db   240  CATATAAAAAGGAGTGGAAAACCATTGCTAATGATGTTGTTCTTAGCAACAACAAAGCAAGA   299
Qy   753  gagggtaaactacagccacactgaattccaagggtggcacctttacgatctccaatttagga 812
Db   300  GAGGGTAAACTACAGCCACATGAATTCCAGGGTGGCACTTTTACGATCTCCAATTTAGGA   359
Qy   813  atgttt-ggaattaaacaatttctctgctatattaccccaccctccaagcactgtattttggc 871
Db   360  ATGTTTGGAATTAAGAATTTCTCTGCTATTATTACCACCACCTCAAGCACATGTATTTTGGC   419
Qy   872  aattggtgcttcaggagagataaactggtccctcgcagataaatgaaaaaagggttttgatgtggc 931
Db   420  AATTGGTGCTTCAGGAGGATAAACTGGTCCCCGCAGATAAATGAAAAAAGGGTTTTGATGTGGC   479
Qy   932  tagcatgagtgtctgttcacactcagttgtgatcaccgggtgtggtggatggagcaagttggagc 991
Db   480  TAGCATGAGTGTCTGTTCACACTCAGTTGTCTGATCACCGGGTGTGGTGGATGGAGCAGTTGGAGC   539
Qy   992  ccagtggctt 1001
          |||||||||||
Db   540  CCAGTGGCTT 549
```

Right section

```
ACCESSION     BE850747
VERSION       BE850747.1   GI:10309106
KEYWORDS      EST.
SOURCE        Mus musculus
  ORGANISM    house mouse.
              Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1  (bases 1 to 654)
  AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
  JOURNAL     Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:1389612
              MGC:1389612
              Seq primer: -40RP from Gibco
              High quality sequence stop: 405.

FEATURES          Location/Qualifiers
      source    1. .654
                  /organism="Mus musculus"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:3470252"
                  /clone_lib="Soares_thymus_2NbMT"
                  /sex="male"
                  /tissue_type="Thymus"
                  /dev_stage="4 weeks"
                  /lab_host="DH10B"
                  /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTACCAATCTGAAGTGGGACGCCGCGGCCGCTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. RNA
                      provided by Dr. Bertrand Jordan. Library went through two
                      rounds of normalization, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."

BASE COUNT    182 a    130 c    159 g    182 t    1 others
ORIGIN
```

```
Qy   372  atcccaatcaagcaacattcgtcgggttattgcacagccgattaatgcaatcaaagcaaacc 431
Db     1  ATCCCCATACCAGCAACATTCGTCGCGAGTGATTGCCGAGAGGCGTCAGCTCGAAGCAGACT    60

Query Match       51.5%;  Score 515.4;  DB 10;  Length 654;
Best Local Similarity   88.6%;  Pred. No. 5e-123;
Matches  558;  Conservative  0;  Mismatches  72;  Indels  0;  Gaps  0;
```

```
Qy   372  atcccaatcaagcaacattcgtcgggttattgcacagccgattaatgcaatcaaagcaaacc 431
Db     1  ATCCCCATACCAGCAACATTCGTCGCGAGTGATTGCCGAGAGGCGTCAGCTCGAAGCAGACT    60
Qy   432  atacctcattattaccttctatcgatcaagataaaatatgggaaaggaagagaagtttgtttgttacggaa 491
Db    61  ATACCTCATTATTACCTTCTGTCGATGTAAAATATGGGAGAGGGTGCTGTTGGTTACGGAAG   120
Qy   492  gaacttaataagatattagaaggagaagaagcaaaaatttctgtcaatgacttcatcataaaa 551
Db   121  GAACTTAATAAGATATTGAAGGAGGAAAAGGAAAATCTCCGTCAACGACTTCATTATATAAAG   180
Qy   552  gcttcagctttggcatgttataaaagttcccgaagcaaatttcttcttgggatggacacagtt 611
Db   181  GCTTCAGCTTTGGCCTGTTCGTGTAAAAGTTCCCGAAGCAAAACTGTCTTGATTGGACACAGTT   240
Qy   612  ataagacaaaatcatgttgatgtcagtgttgcggtcagtactcctgcaggactcatc 671
Db   241  ATAAGGCAAAATCATGTGGTTGACGTCAGTGTTGCCTGTCAGTACCCCTGCAGGACTTATC   300
Qy   672  acacctattgttactgcacatataaaaggagtggaaaccattgctaatgatgttgtt 731
Db   301  ACCCCTATTGTGTTTAATGCACACATAAAAGGACTGGAAACCATTGCTAGTGATGTTGTT   360
Qy   732  tctttagcaacaaagcaagagagggtaaactacagccacactgaattccaagggtggcact 791
```

```
RESULT 14
LOCUS        BE850747    654 bp    mRNA    linear    EST 26-SEP-2000
DEFINITION   ux01e11.y1 Soares_thymus_2NbMT Mus musculus cDNA clone
             IMAGE:3470252 5' similar to SW:ODP2_RAT P08461 DIHYDROLIPOAMIDE
             ACETYLTRANSFERASE COMPONENT OF PYRUVATE DEHYDROGENASE COMPLEX ;,
             mRNA sequence.
```

```
LOCUS       BI658188               774 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603285296F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5329774 5',
            mRNA sequence.
ACCESSION   BI658188
VERSION     BI658188.1  GI:15572424
KEYWORDS    EST.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11836 row: o column: 23
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1..774
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5329774"
                     /clone_lib="NIH_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert 2.5 kb. Library constructed by Life
                     Technologies, catalog # 12018-016. Investigators providing
                     samples: Lothar Hennighausen/Priscilla Furth, NIH
                     Reference for transgenic model: Li et al., Cell Growth and
                     Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
                     Library."
BASE COUNT      209 a    205 c    188 g    172 t
ORIGIN
```

```
Qy      1 atatatcacgcattgtgactataggcccaaccgaagtaacagatttaaaccacaagtgc  60
            ||||||| |||||||||||| | || |||||||||||||||||||||||||||||||||
Db     38 ATATAGCACGCATTTGCAGACTACAGGCCACTACAGAGAAGTGACCACAGCTTAAAGCCACAGGCAG  97

Qy     61 caccaacctaccaccccccggttggccgttcctccacaactcccagctctttagcctcta  120
            ||||||| || || |||||| |||| ||||||| |||||||||||||||| |||
Db     98 CACCAACCTGCCCACCCCCCGGTGGCCGTCGTTCCTCCACCTCCCAGCGCCCCAGCCTTA  157

Qy    121 cacccttcagcaccctgcccagctcactcctgctcggaccaaagtgaaaaggggttcttgttagc  180
            ||||| |||||||| ||||||| |||||||||||||| ||||| ||||||||||||||
Db    158 CCCC--------CTCAGCGTCCTCCTGTCCTGGACCAAAAGGGAAGGGTGTTCGTTAGTC  205

Qy    181 ctcttgcaaagaagttggcagttagagaaagggaggttgatcttacacagtaaaaggacaag  240
            |||||| |||||||| |||||| |||||| |||||| |||||||||||||||||||| ||
Db    206 CTCTTGCGAAGAACGTTGGCAGCAGAGAAGGAGGATTGACCAAAGGGACCACAAGTTAAAGGGACGG  265

Qy    241 gaccagatggtagaatcaccaagaagtgatatcgactctttgcctagtgtctctacaggtg  300
            ||||||||||||||||||||| |||||||||||| |||||||| || |||||||||||||
Db    266 GACCAGAAGGCAGAAATCAATCACAAGAAGGACATTGAACTCTTTTGTGCCTTCTAAGGCTGTC  325

Qy    301 ctgtccccggcagctgttgttgtcctcccacaggtcctggaatggcaccagttcctacaggtg  360
            |||| ||||||| |||||| ||||||| |||| |||||||| |||||||||||| |||||||
Db    326 CTGCTGCCAGCCAGGTGCCATGGCGGTCCAAGAGTGGGCACCAGCAGCTCCTGCAGGTG  385

Qy    361 tcttcacagatatcccaatcagcaacatccgtcgggttattgcacagcaggaattaatgcaat  420
            |||||| ||||||||||| |||||||||||| ||| ||||| |||| |||||||||||||||
Db    386 TCTTCACAGACATCCCATCAGCAACCATCCGTCGAGTGATTGCCGCAGAGGCTCAATGCCAGT  445

Qy    421 -caaagcaaaccatcacctcattattacccttctcatatcgatcgtaaatatgggagaagtttg  479
             |||||||| |||||||||| |||||||||||||||||| ||||| | |||||||||||| |||
Db    446 CCGAAGCACGACTAACCTCAATTATTACCTTTCTCTGTCGGATGTAAATATGGGAGGTGCTG  505

Qy    480 ttggtacggaaagaaacttaataagatattagaagggagaa---agcaaaaatttctgtcaa  535
            ||||| || |||||||||||||| ||||||| |||| ||||   ||||||||| | ||| ||
Db    506 CTGGTACGGAAGGAAACTTACTAAGGATGCTTGACGGGACACAGCGGAAACATCTCCGTCAA  565

Qy    536 tgacttcatcataacagccttcaagtcttttggcatgtgtttaaaagttcccgaagcaaatttctc  595
            ||||||||||| |||||| ||||||| ||||||||| ||| | ||||||||||| |||||||
Db    566 CGACTTCATTATAAAGGCTTCAGCTTTGGCCTGTGTCTGAAAGTTCCCGAAGCAAACTCGTC  625

Qy    596 ttggatggacacagttataagacaaaatcatgttgttgatgtcagtgttgcggtcagtac  655
            ||||||||||||| |||||||||||||||||||| |||||||||||||| ||| ||||||
Db    626 TTGGATGGACACAGTTATAAGGCAAAAATCATGTCGGTTGACGTCAGTGTAGCCTGTCAGTAC  685

Qy    656 tcctgcaggagctcatccacacactattgtgtttaatgcacatataaaagggagtggaaaccat  715
            ||||||||| ||| ||||| |||| ||||| ||||||||||| |||||||||||| ||||||
Db    686 CCCTGCAGGAGACTTATCACCCCTATAGTGTATAAATGCACACAATAAAAGGACTGGAAACCAT  745

Qy    716 tgctaatgatgttgttctttagcaaacca  744
            ||||| |||||| ||| |||||||||||
Db    746 TGCTAGTGTGATGTAGTAAACTTTAGCCTGCA  774
```

```
Qy    361 TCTTTAGCCTCTCCAAACGGCGAGAGGGTAAACTTCAGCCTCATGAGTCAGGGTGGGAACA
            |||||| |||| | || |||||||||| || |||||||||||| |||||||||||||| ||
Db
```

```
Qy    421 TTTACAATCTCCAACTTAGGGATGTTTGGAATTAAGAATTTCTCTGCAATTATTAATCCA  480
            |||| | |||||||| ||| ||||||||| |||||| ||||||||||||||||||||| |
Db    421

Qy    481 CCTCAGGCATGTATTCTTGGCAATCGGTGCTTCAGAGGATAAACTGATCCCACCAGATAAT  540
Db    541

Qy    792 tttacgaatcccaaattaggaaatgttggatta aaagaaatttctctgctattaccca
            ||||| || ||||||||||| |||||||||| |||||||||| | |||| |||||||
Db    851

Qy    852 cctcaagcatgtattttgtgcaattggtgcttcaagaggaggata aaactgtccctgcagataat  911
Db    912 gaaaaggggtttgatggtgctagcatggcatgatagtctgttcacactccagtcttgtgatcaccgggtg  971

Qy    972 gtggatgggagcaagttggagcccaagtggctt  1001
            ||| ||||||||| |||||| | |||||||||
Db    601 GTGGATGGGAGGCAGAGGATGGGAGCCCCANTGGCTT  630
```

```
Query Match      50.8%;  Score 509;  DB 10;  Length 774;
Best Local Similarity    83.0%;  Pred. No. 2.5e-121;
Matches  622;  Conservative    0;  Mismatches  110;  Indels  17;  Gaps    3;
```

```
Search completed: August 13, 2002, 18:59:28
Job time: 5135 sec
```